THE UNIVERSITY OF ALBERTA

A STUDY OF THE VALIDITY OF USING THE LORGE-THORNDIKE AND RAVEN'S STANDARD PROGRESSIVE MATRICES TESTS IN THE IDENTIFICATION OF UNDERACHIEVERS

by

(C)

GREG FEDUN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1981

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Greg Fedun

TITLE OF THESIS A Study of the Validity of Using the Lorge-Thorndike and Raven's Standard Progressive Matrices in the Identification of Underachievers

DEGREE FOR WHICH THESIS WAS PRESENTED M. Ed.

YEAR THIS DEGREE IS GRANTED 1981

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed) Greg Fedun

PERMANENT ADDRESS:

Box 154

Grand Centre

DATED August 1981

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled A Study of the Validity of Using the Lorge-Thorndike and Raven's Standard Progressive Matrices Tests in the Identification of Underachievers submitted by Greg Fedun in partial fulfilment of the requirement for the degree of Master of Education in Educational Psychology.

Supervisor

.....

.....

Date 14 August 1981

Abstract

This study was undertaken to determine whether beta errors occur when group intelligence measures are used in the identification of underachievers, in a population of students with varied language backgrounds. The Lorge-Thorndike, as it was in use in the schools at the time, was one of the tests investigated. Raven's Standard Progressive Matrices Test was also administered to determine its adequacy as an aide in identifying underachievers. As the WISC-R is used within the schools to make placement decisions it was used as the intelligence criterion measure with which both group intelligence tests were compared.

The procedure involved comparing each individual's intelligence score to his achievement test score (on the Canadian Tests of Basic Skills) using standard score units. If there was a discrepancy of one standard deviation or more in favor of the intelligence score, the student was classified as an underachiever. Through a series of crosstabulations, the students identified as underachievers on the WISC-R were compared to those identified as underachievers on each group test.

The results revealed that numerous beta errors occurred when the Lorge-Thorndike was used following this procedure. When the Standard Progressive Matrices Test was used, fewer beta errors occurred. The results also indicated that the Raven's measured mental abilities were more evenly distributed across the language background groups, and that these abilities were not identical to those measured by the other intelligence measures.

An exploration appears warranted to find a more recent and adequate group intelligence instrument to supplant the Lorge-Thorndike. It also appears that consideration should be given to incorporating the Standard Progressive Matrices Test into the assessment procedure for individuals who are minority group members, allowing for a different perspective of their mental abilities.

ACKNOWLEDGEMENTS

There are individuals, without whose support this project would not have been completed.

The author wishes to express indebtedness and gratitude to his parents, Leo and Anne Fedun.

At the same time, he wishes to acknowledge the support and encouragement of the following people:

Dr. G. Fitzsimmons, for his interest, direction and encouragement, as a supervisor and friend.

Dr. H. Janzen and Dr. A. Olson for their continued confidence and patience.

Corine Strickfaden for her valuable assistance in editing and organizing the manuscript.

Sandy Shortell for her exceptional consideration, and time spent, in typing the manuscript.

To the Lakeland Catholic School Division, teachers, administration and board, for their cooperation and support of this study.

To my sons, Simon and Kristopher, and the rest of my family, for their understanding and prayers.

TABLE OF CONTENTS

CHAPTER		PAGE
LIST OF TABLES		ix
1.	INTRODUCTION TO THE STUDY	1
	Limitations of the Study	4
2.	REVIEW OF THE LITERATURE	7
	Theories of Intelligence	7
	Development of Intellectual Assessment Instruments	11
	Nature of Intelligence Tests	13
	Minority Group Bias in Intelligence Testing	20
	Uses and Consequences of Testing	28
	Implications of the Review of the Literature for this Study	32
3.	METHOD	36
	Population	36
	Data Gathering Procedure	36
	The Measurement Instruments	38
	The Canadian Tests of Basic Skills	39
	The Wechsler Intelligence Scale for Children - Revised	40
	The Canadian Lorge-Thorndike Intelligence Tests	47
	Raven's Standard Progressive Matrices	44
	Data Analysis Procedure	45
4.	ANALYSIS OF THE RESULTS	48
	Means and Standard Deviations of Data Grouped by Grade	48
	Standard Score Analysis	51
	Correlations	53
	Crosstabulations	55

PAGE

5. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS..... 61

Discussion of the Results 61

Conclusions 66

Recommendations 67

Suggestions for Further Research 69

REFERENCES 71

APPENDICES 84

APPENDIX A - Distribution of Scores for the Various Group Measures by Grade and by Group. 85

APPENDIX B - Scattergrams Indicating the Extent of Agreement Between the Raven's and L-T(FS) for Each Group and Total Group..... 90

APPENDIX C - Scattergrams Indicating the Extent of Agreement Among the Raven's, L-T(FS), and WISC-R (FS) for Total Group 97

LIST OF TABLES

TABLE	DESCRIPTION	PAGE
1.	Table of Abbreviations and Symbols Used in This Study.	47
2.	Means and Standard Deviations of Raw Scores by Grades.	49
3.	Standard Score Means, Standard Deviations and ANOVA by Groups	52
4.	Correlations Between Measures	54
5.	Crosstabulations of Classified Students Among the Raven's, L-T (V), L-T (NV) and L-T (FS) for Groups and Total Group	56
6.	Crosstabulations of the Classified Students Among the Raven's, L-T (FS) and WISC-R (FS) for Groups and Total Group	58
7.	Crosstabulations of Classified Students on the L-T (V) and (NV), WISC-R (V) and (P), and the Raven's for the Total Group	59

CHAPTER 1

Introduction to the Study

There is a general consensus of opinion that the ultimate aim of testing in the schools is to enhance the education of the individual child. There appears, however, to be little agreement as to whether this aid is in fact achieved through the present testing programs of our schools.

Each school district, and sometimes each school, has its own special blend of assessment materials. These include tests that are: group or individually administered, norm or criterion referenced, bought or teacher made, and ability or achievement oriented.

The stated purposes for administering the tests are varied. The reasons given for testing include: establishing the present level of intellectual or academic functioning, determining the amount of change that has taken place over a period of time, predicting future levels of achievement, determining patterns of relative strengths and weaknesses and identifying students' specific needs.

The subject of this thesis is an investigation into the use of the Lorge-Thorndike in the identification of underachievers in the Grand Centre - Cold Lake area. In the three Catholic schools involved in this study, the Lorge-Thorndike is administered in Grades 4 and 7 annually. Those students who are identified as average or above average in intellectual ability by the Lorge-Thorndike, and who score in the below average range on the Canadian Test of Basic Skills, a standardized achievement test, are tentatively identified as underachievers.

A significant portion of the students who live in this area come from homes where the parents speak another language, primarily Cree, French or Chipewyan. For these students, the Lorge-Thorndike may be an inappropriate instrument. The Lorge-Thorndike has a verbal and nonverbal section with a significant emphasis on reading ability and vocabulary development. Due to this verbal loading, the students from homes where another language is spoken may be at a disadvantage. This may result in spuriously low estimates of intellectual ability for them.

In using the Lorge-Thorndike for the purpose of identifying underachievers in this population, beta errors, that is, false negative errors could result. In this study beta errors occur when an assessment instrument underestimates the intellectual ability of an individual. Under these circumstances no significant difference between the achievement level and the intellectual ability is evident.

A possible outgrowth of this type of error is that these students would not be recognized as underachievers. They would then be missed as candidates scheduled for more extensive individualized assessments. There is also the possibility, that teachers' expectations and objectives might be at an inappropriate level, in light of inaccurate intellectual ability information.

The administrators in the schools as well as the teachers reached a consensus on an operational definition of an underachiever. The operational definition arrived at was the following: A student will be considered a possible underachiever if his/her standard score on either the Verbal or Nonverbal component of an intelligence test exceeds his/her standard score on the Canadian Test of Basic Skills

(C.T.B.S.) by one standard deviation or more. The classification system of identifying underachievers on the basis of a discrepancy between intelligence and achievement test results has been used by others such as Ellis (1969). Though the present definition does not incorporate the standard error of measurement of each assessment instrument, the possibility of errors in identifying students as underachievers who prove not to be underachievers, is judged by the author, as preferable to errors in the opposite direction.

Subsequently, and for consistency, the operational definition of an overachiever is a student whose standard score on the C.T.B.S. exceeds his/her standard score on an intelligence test by one standard deviation or more.

At this point the author wishes to underline that the operational definition does not connote value judgments that are at times associated with the term underachiever. This term has at times been synonymous with terms such as lazy student, or behavior problem in the class. Also, the terms underachiever and overachiever are both misnomers to some extent. Both terms tend to imply that an appropriate level of achievement exists for each student. This implication is not intended by the author. The classifications are used primarily for consistency since they have been traditionally used to identify students, and are thus readily understood.

On the basis of the operational definition, students identified as underachievers and overachievers, when the Lorge-Thorndike verbal IQ test is used, will be compared to the students identified when the Raven's is used as the intelligence test. This same procedure will be followed in comparing the Lorge-Thorndike nonverbal test and the Raven's

results. A final comparison will be made using the WISC-R as the criterion reference. Those identified under the overachievers when the WISC-R is used, will be compared to those identified by the group tests.

The significance of this study lies in the fact that the validity of the Lorge-Thorndike, for students in this region, has not been investigated. Validity in this context is defined by N.Grunbind as follows:

> "validity, in general, refers to the extent in which the results of an evaluation procedure serve the particular uses for which they are intended" (1976, p.26).

Since this validity has not been established, gifted students who are achieving at an average level, or students of average intellectual ability who have a low achievement level, might have been overlooked. This would be a distinct likelihood if the Lorge-Thorndike proved to be an inadequate or inappropriate assessment instrument for these students.

Limitations of the Study

There are three specific areas that are beyond the scope of this study. The first is the use of the WISC-R as the criterion for judging the reliability of the group intelligence tests. The Alberta Department of Education has to this point insisted that each student placed into a non-average intellectual category, such as Gifted, Educably Mentally Handicapped, or Learning Disabled, undergo an individual assessment. The assessment is to include a measure of intelligence such as the WISC-R or Stanford-Binet. School grants, special funds and other forms of assistance are based on the intellectual assessment results. It is beyond the scope of this study to delve into the merits of this policy and procedure.

The possible factors contributing to a significant difference between intellectual ability, and level of academic achievement will not be investigated. However, within the three schools involved, a variety of techniques are used to determine possible negative influences on a student's progress. All students who are possible underachievers are individually assessed with a WISC-R. Factors of hearing, vision and physical health are eliminated as possible contributing negative influences. In order to prescribe an appropriate remedial program for the student, data is collected in one or more of the following ways. A behavioral observation is done. An examination of the individual's pattern of relative strengths and weaknesses is performed. Diagnostic tests of perception, co-ordination and core subject skills are administered. In-depth study of the student's history is done using sources such as: the cumulative record of the student, student and parent conferences and information supplied by the teachers.

Further, this study will not discuss the desirability, or value, of expending time, energy, and finances in identifying underachievers, or in the intellectual categorization of students. This study is not intended to determine whether gifted students, for example, should receive more attention than students with learning disabilities, or those that need some form of psychological support or guidance.

Although this initial investigation may indirectly shed light on these unresolved issues, to include them would expand the study beyond its present mandate. The author is also aware that these issues are the primary focus of numerous intensive investigations and discussions, such as these carried out by the Alberta Teachers Association.

The final implications of this study, to determine the validity of using the Lorge-Thorndike Intelligence Test in the identification of underachievers of students with varied language backgrounds, have not been established. Thus, the need for an investigation such as this is substantiated.

CHAPTER 2

Review of the Literature

> "Now I.Q. testing is outlawed in San Francisco, personnel selection tests are declared illegal unless directly related to employment, group intelligence measures are banned in the New York City schools, a whole profession which has distinguished itself from psychiatry primarily because its practitioners can test has been declared moribund, and school psychologists in Boston have been declared incompetent. In the last ten years, what was once a silk purse has been transformed into a sow's ear." (Bersoff, 1973, p.982)

Bersoff's rather colorful and dramatic statements reflect the fact that psychological assessment and intelligence testing in particular has generated a great deal of controversy in recent years. Salvia and Ysseldyke (1981) suggest that the controversy centres around three issues. The first is questioning the nature of intelligence testing, that is, the premises under which they were constructed, and what they measure. Then, questioning the appropriateness of the tests for the population it evaluates, and finally, questioning the social consequences of the testing.

In order to bring these issues into perspective, this chapter will briefly review the theories of intelligence underlying intellectual assessment and present an overview of some of the major historic steps in the development of intelligence testing. The literature focussing on the major areas of contraversy will also be reviewed. Implications of the review, and their relationship to this study, will conclude the chapter.

Theories of Intelligence

According to Sattler (1974), theories of intelligence developed in different directions, depending on the focus of the particular proponent. At a symposium in 1921, thirteen psychologists, though much in agreement, presented thirteen different views of the nature of intelligence

(Terman, 1921). Some psychologists have used factor analysis to arrive at, and support their theories. C.E. Spearman was an early proponent of this method. He arrived at a two factor theory. One factor consisted of the actual ability being measured on any given instrument. The other factor he named a general factor of "g", which is shared by every specific factor to a greater or lesser degree (Spearman, 1932).

L.L.Thurstone's work led him to propose that there were seven important groups of factors, which he called primary mental abilities. On this basis he constructed the Primary Mental Abilities Test (1965). Later Thurstone postulated a second-order "g" factor, when his primary factors were found to correlate moderately among themselves.

In attempting to synthesize the work of Spearman and Thurstone, P.E. Vernon (1961), formulated a theory where "g" was of central importance. Next in importance were two factors, Verbal-Educational and Practical-Mechanical-Spatial. These factors were broken down to minor group factors, and at the last level, broken down to specific factors peculiar to certain tests.

J.P. Guilford (1967) a prominent theorist in factor analysis, proposed a three dimensional model called the Structure-of-Intellect (SI). The SI model has a mental operations dimension (five factors), a content dimension (four factors), and a product dimension (six factors). He therefore postulates the existence of 120 major factors. Guilford's model has been criticized, as proof is lacking of a central feature, positively correlating the factors, has not been demonstrated or provided (McClemar, 1964). Guilford, however, feels that his model more accurately reflects the interrelationship of intellectual abilities,

than models relying on one or two factors (Guilford, 1971).

R.L. Thorndike's premise was that intelligence was composed of numerous elements that formed clusters. He identified three such clusters, concrete intelligence, dealing with things, social intelligence, dealing with people, and abstract intelligence, dealing with symbolic manipulations, (Thorndike and Hagen, 1977). It is the abstract intelligence that Thorndike is attempting to assess with the Lorge-Thorndike Intelligence Test (Lorge et. al., 1967).

Some psychologists have incorporated genetics as a feature of their definitions of intelligence. One of the first to do so was Hebb (1966). He suggested that there were two elements that were subsumed in the term "intelligence". He labelled these elements intelligence A and intelligence B. According to Hebb, intelligence A refers to the innate, inherited intelligence that cannot be measured directly. Intelligence B refers to the intelligence arising from the interaction of the individual with his environment. Thus, intelligence A always contributes to intelligence B.

Genetics is a significant factor in J. Piaget's theory of intelligence as well. According to Piaget, intelligence develops out of the interaction between the processes of assimilation, responding to inner or biological urgings, and accommodation, responding to the environment. Higher order rational mental processes develop as the individual's mental processes become more independent of both the inner and environmental promptings (Greburg and Opper, 1979).

It appears that psychologists are reaching more of a consensus as to the nature of intelligence. As previously mentioned, Vernon's theory attempts to reconcile Spearman's and Thurstone's views on the

importance of the "g" factor. J.M. Hunt (1961) points out that Hebb's intelligences A and B correspond closely to Piaget's processes of assimilation and accommodation. Vernon (1969) points out that Cattell's theory closely parallels Hebb's. Cattell (1963) postulated that there are two types of intelligence, fluid, a basic inherent ability to learn, and crystallized, resulting from fluid intelligence interacting with the culture. Vernon points out that fluid intelligence approximates Hebb's intelligence A and crystallized approximates intelligence B.

Another psychologist who has attempted to synthesize different theories into ine is A.R. Jensen. He agrees with the factor analytic theorists who postulate a "g" factor. Jensen feels that there is a general intelligence that is tapped to greater and lesser degrees by various intelligence tests (Jensen, 1970). He also affirmed the position of psychologists who feel genetic factors are significant in the development of intellectual ability (Jensen, 1969). Jensen also postulates the existence of hierarchical learning abilities, based on his studies of blacks, whites and Mexican Americans in California (1973), and blacks and whites in Georgia (1977). According to Jense, each child has associative learning ability to about the same degree. This allows each person to function in every day life. The higher ability, the conceptual learning ability, is not as evently distributed genetically and a poor impoverished environmental background could result in a minimal development of this ability. Jensen concludes that this initial cultural deprivation influence on conceptual learning development accounts for the increasingly poor performance of some students in their school subjects. He refers to this as the cumulative deficit hypothesis (Jensen, 1974).

From the brief review of the various theories, it would suggest that Sattler is correct when he states:

> "Theories of intelligence are beginning to show a coalescing of views, stressing the importance of both innate and developmental influences. Intelligence is viewed as being a central, "fluid" genetically determined basis ability which is modified by experience." (Sattler, 1974, p.15)

It is also evident ahat total consensus between psychologists is nowhere near. An example of this is Guilford's recent defence of his SI model. He contends that the research of Horn and Cattell (1966) into fluid and crystallized intelligence is actually identifying second order factors in his SI model (Guilford, 1980).

Development of Intellectual Assessment Instruments

One of the early contributions to the field of intellectual assessment was made by Sir Francis Galton. In attempting to assess intelligence, he focussed on the ability of an individual to make fine sensory discriminations. As a result of this focus, his research met with limited success, though he did contribute to statistics gathering, and analysis techniques. (Achurst, 1970). At approximately the same time, K. Pearson was working on the development of his correlation coefficient, and J.M.Cattell was developing statistical procedures used in the evaluation and application of measures.

In the early 1900's, A. Binet was commissioned to develop a test (Achurst, 1970) that would identify students who would not benefit from regular schooling. Working with V. Henri, and later T. Simon, he focussed his attention on the higher mental processes. He based his test selection on its ability to discriminate between younger and older children. Sorting these tests into different age levels led to the development of a scale. The scale could be applied and scored in a standard manner and was known as the Binet-Simon scale. Scores on this

scale were better referred to as mental ages (Tuddenham, 1962).

In America, L.M. Terman revised the Binet-Simon scale extensively, and developed the ratio I.Q.. The I.Q. is arrived at by dividing the mental age by the chronological age, and multiplying by 100. Since these scores were not comparable across age groups and different tests, a Deviation I.Q. (D.I.Q.) was developed. G. Thomson was instrumental in developing the I.Q. which converted scores on a test onto a standard score with a mena of 100 (Tuddenham, 1962).

Group intelligence tests were initially experimented with by A.S. Otis in American and C. Burt in England. With the development of the Army Alpha and Beta tests, group tests were used to screen military personnel, and to assign them to various positions. As they could be quickly administered and scored, group test use was extended to the civil service, colleges and to businesses (Goslin, 1963).

Wechsler viewed intelligence, as being global in nature, with various factors entering into its composition. He studied the various tests available at the time, in the thirties, and drew eleven subtests, from tests such as the Stanford-Binet, and Army Alpha. He didn't rank the subtests in any hierarchical order, as he felt the overall I.Q. obtained represented a measure of "g" or general mental ability (Wechsler, 1958).

Another development of testing was the creation of culture fair tests. Psychologists who agreed that a general factor was present in all intellectual assessments, attempted to create instruments that would measure this factor almost exclusively. The rationale was that this would allow assessment of intelligence across cultures and environments. Two well known tests of this nature are Cattell's Culture Fair Intelligence Test (1959), and Raven's Progressive Matrices Test (1960).

At present, other avenues are being explored in order to determine the nature of intelligence, and how best to assess it. Piaget (1978) felt that the psychologists probing semantics and language structure may be able to make a useful contribution in this area. One theory receiving serious attention is Vygotsky's (1962). He proposed that the language experiences of an individual structure the development of his intellect. That is. logical thought processes are created through the internalization of speech. Peal and Lambert (1962), in their study of the intellectual development of bilinguals, feel that Vygotsky's theory merits further research. They feel it can contribute to the understanding of mental abilities associated with the reasoning processes.

Although this brief review has been presented sequentially, the development of intelligence assessment instruments has not been a linear progression. Rather, intelligence testing was, and is, a controversial, and divergent topic.

Nature of Intelligence Tests

The history of intellectual assessment reveals that intelligence tests have been misused, and their results inappropriately interpreted. Sarasan and Doris (1969), and Kawin (1975), cite a veritable littany of offences. Some of the more serious offences included classifying the poor as feebleminded, condemning different nationalities and races on the basis of appearance. These occurrances have not contributed to the understanding or acceptance of intelligence testing. Aspects of intelligence tests that need clarification include the following:

A Clear Indication of What the Intelligence Test Measures:

They do not measure innate potential, although earlier psychologists may have thought they were measuring potential.

Many psychologists such as Jensen (1970), Clarizio (1979) and Flaugher (1979) would appear to agree with Thorndike's statement that:

> "There is no question that during the early years of testing there was a good deal of enthusiasm and naivete in the use of the tests, in (thinking about) what they measured - the concept of somehow tapping native ability. I think that nobody at present would contend this is the case. Everybody would acknowledge that we would have no way of directly measuring native ability. What we measure is the developed abilities the individual possesses at a given level and point in time." (Thorndike, 1978, p.18)

Jensen (1977) seems to represent the current thinking regarding what tests measure. He states that besides measuring a general factor, to some degree, each test measures a sampling of intellectual ability, out of a broad spectrum of abilities. Vernon (1979) goes further, by suggesting that an intelligence C should be added to Hebb's intelligences A and B. This, he posits,would clarify the issue of whether a person was referring to genetically based intelligence (A), the more general construct of intelligence (B), or the result obtained on a specific test (C). In agreeing that a test measures specific abilities, Thorndike (1959) and Salvia and Ysseldyke (1981) strongly recommend that every test should be scrutinized as to what abilities it proports to measure, before it is used.

Criticism of intelligence testing has arisen regarding discussion of tests in a manner suggesting that a fixed, immutable trait of an individual is being measured, although it is the individual's specific abilities at a given time that are being assessed. Both Anastasi (1970) and Tryon (1979) have criticized psychologists such as Thurston, Jensen and Thorndike for making associations between a perceived test performance, and a presumed mental function. Tryon is also critical of Mercer, who has developed a system that is meant to evaluate the educational potential of minority group students. Tryon feels that

Mercer's claim of being able to estimate potential has no validity in the light of present scientific methodology (Mercer, 1977). Bersoff (1973) points out that the psychologists themselves contribute to the confusion between ability and trait. He feels that the practice of describing even very simple ability measures in more general terms may add more authority to the instrument than it warrants, and add to the confusion, or misunderstanding of what it actually measures.

Fuller (1977) is critical of test result reporting. He questions the adequacy of a single number representing the varied mental abilities measured in an assessment. According to Fuller, arriving at one number interferes with our understanding of an individual as it lumps all of that person's abilities into one amorphous whole. He feels that the specific strengths and weaknesses that individualize a person should be reported as such, without any attempt to group them into a meaningless number.

In order to avoid these criticisms, Salvia and Ysseldyke (1981) state most strongly that creaters of intelligence tests clearly indicate the reasoning and rationale that form the basis of the intelligence test, and what specific abilities it is intended to measure. Further, they should provide reliability and validity data, based on well constructed research studies.

The Stability of I.Q. Scores:

Bloom (1964) attempted to establish the stability of intelligence over time, by reviewing, and reanalyzing research statistics from previous studies, as well as performing his own longitudinal investigation. His intent was to establish what percentage of an individual's I.Q. could be accounted for at a geven age. He concluded that infant intelligence was not fixed and was affected by environmental

factors. Bloom felt that environmental factors were much less significant after the age of thirteen. He also concluded that individual assessments were much more accurate in estimating intelligence at every age level than were group tests/ These results were tentative, as Bloom faced the same obstacles accompanying most longitudinal studies, the attrition of subjects, and the intrusion of extraneous variables.

Skodak and Skeels (in Jensen, 1973) in their study in 1945, indicated that babies given up for adoption by unwed mothers, had I.Q.'s significantly superior to their mothers. They concluded that the enriched environment of the new home accounted for the difference. Hunt (1961) also felt that the environmental influences were the significant factors for the lower I.Q. scores of deprived and minority group children. As a result, Hunt was a firm supporter of Project Head Start, a United States government sponsored program designed to assist the intellectual development of environmentally impoverished children.

Rosenthal and Jacobson's (1968) study also suggested that environmental influences could affect the intellectual development of a student. Their study consisted of giving teachers the impression that certain students could, or would soon, function at a higher intellectual level. According to the researchers, the students identified as "spurters" made significant I.Q. gains. Since these students were actually identified at random by the psychologists, they concluded that the reason for these gains was the change in the teacher's expectations.

Rosenthal's and Jacobson's findings have met with criticism from their colleagues. Thorndike, reported in Cronbach, (1975), felt that the results should never have reached the publishing stage.

Cronbach, in the same article, finds fault with the design of the study, the way the data was analyzed, and with the fact that results contradictory to their conclusions, received minimal attention. Though Rosenthal defended his research and conclusions (1973), Sattler (1974) reports that the attempts to replicate Rosenthal's study have been unable to arrive at similar conclusions.

Jensen (1973) strongly criticized the Skodak and Skeels study. He concluded that the researchers had overinterpreted incomplete data. He also suggested, that the results indicating I.Q. differences between mother and child, were still compatible with his belief in the inheritability of intelligence. He has always affirmed, that environmental factors can affect a person's intellectual development to a significant extent.

Relative Advantages/Disadvantages of Group and Individual Tests:

Anastasi (1976) is quite comprehensive in her coverage of the advantages of group tests. She suggests that they are a saving in both administration and scoring time, due to the fact that they usually have a multiple choice format. They effectively standardize administration procedure and do away with the need for specifically trained experts. Group tests are also adequate gross screening devices, and are convenient to use when making between group, across group, and longitudinal group comparisons. A weakness of individual assessment, according to Williams and Kirkland (1971), is that if the examiner and examinee are of different ethnic origins, or different backgrounds, both biasness, and communication difficulties can arise. Bersoff (1973) questions the validity of an assessment result that is obtained in an optimum setting, in a one-to-one situation. He suggests that since the

classroom and testing environments are dissimilar, the carry over of the assessment results to the classroom will be minimal. Presumably, the group assessment setting resembles the regular educational setting of the child, thus, the group test results are more valid. On the other hand, there are critics of group testing. Thorndike and Hagen (1977) point out that group tests allow for minimal contact between the administrator and student, making it difficult for the administrator to monitor each student's progress. Further, the student has no room to record alternate responses to the questions, and the timed tests do not make allowance for students who have different work rates. There is little consideration of a student's level of preparedness to perform, and overall, these tests are of little value in clinical situations.

Sattler (1974) points out that individual tests are superior to group tests, in that they are more adequate predictors of achievement, and they yield a more useful picture of cognitive development. Anastasi (1976) and Salvia and Ysseldyke (1981) agree that no placement decision should be made on the basis of a group test, and that one of the prerequisites should be an individual assessment.

The Importance of Intelligence Assessment Results:

The relative importance of an assessment result would be dependent on the perception of the evaluator. Generally, psychologists refer to the meaning of test scores in terms of their statistical significance, or their implications for theories and research. Psychologists such as Vernon (1979), Thorndike and Hagen (1977), and Bryan and Bryan (1978) agree that an intelligence test result is but one of the factors that contribute to the understanding of an individual.

None of them have suggested that intelligence is the single most important factor in the assessment of an individual. A number of indicators point to the fact that this position is not shared by everyone. That individuals view intelligence as very important, is indicated by the reaction that followed Jensen's tentative hypothesis that there was a genetic component to the I.Q. difference between blacks and whites. Cronbach (1975) described this reaction as immediate, emotional, and quite extreme. The reasons for such an unexpectedly strong reactions will be explored in the following sections.

Recent studies indicate that educators tend to attach a great deal of importance to an I.Q. score. In a study by Smith and Knoff (1981), students taking courses in assessment, and psychologists in the schools, were asked to make placement decisions, after being given profiles of students. Both students and psychologists in the field tended to use the I.Q. score as the major support for their placement decisions. In a similar study, Matuzzek and Oakland (1979) found that both teachers and school psychologists felt that the I.Q. result was very important in a placement decision. Barnes (1973) gave teachers information about students' intellectual ability according to a matched distribution schedule. He concluded that when given intellectual assessment information, teachers developed either positive, or negative mental sets towards the students. An example of the importance of I.Q. scores in Alberta, is the fact that the criteria for obtaining funds for children with special needs, is an individual assessment with the resulting I.Q. falling within a given range (Government of Alberta, Department of Education, Category "A" grants document).

Readings dealing with the nature of intelligence tests, including the studies of Jensen, Vernon, and Cattell, lead to the conclusion that the tests measure specific abilities at a given time. Intelligence test results may be influenced by environmental, administrative, and procedural factors. As intelligence test results are but one of the factors that are relevant in the assessment and understanding of an individual, care should be taken to avoid the overemphasis of the significance of this area of testing.

Minority Group Bias in Intelligence Testing Criticisms Regarding Bias in Current Intelligence Tests

The major social issue in intellectual assessment is the controversy over the amount of bias a given assessment instrument contains. The issue involved is the appropriateness of the instrument, for the subjects to whom it is administered. One of the most vocal groups in their criticism of established tests, is the black community. Prior to Jensen's (1969) article, projects such as Head Start were based on the assumption that the difference between the I.Q.'s of blacks and whites was as a result of their generally poorer environmental condition. When Jensen postulated that genetic factors are more significant than the environmental factors, criticism was then focussed on the assessment instrument. R.L.Williams (1970) a leading black psychologist, called for an immediate moritorium of all intelligence testing until less racist tests were available. Macklin and Holman (1976) tended to agree with Williams, adding that in their study of Brooklyn blacks, the test format, item difficulty level and language of the test, in this case the Lorge-Thorndike, was inappropriate for the children it was intending to assess. In their study, Nichols

et.al. (1977) concluded that both black and white children from impoverished backgrounds had difficulty relating to, and communicating with assessors, who tended to be white, and members of the middle class. Lampert (1978) points out that in a legal challenge of testing, the prosecution produced evidence, and witnesses verifying the fact that at times, assessors could not understand what the assessed individual was saying.

Claims of bias have also arisen in the Mexican-American community. One of the most vocal critics of present assessment procedures is J. Mercer. She has done extensive research in the intelligence testing area with Mexican-Americans in California. Mercer (1972) has reached the same conclusions that critics on behalf of blacks had reached, such as Ramirez et. al. (1971) and Zirkel (1972), in their separate studies on the adjustment of Spanish speaking students to regular classrooms. They feel that the totally different language structure of the Spanish language makes it difficult for the students to do well on any tests with a verbal orientation. They also point out that in the development of most intelligence instruments, the Spanish population was not part of the norming group. They feel an accurate assessment of Spanish speaking individuals is not possible with the present assessment instruments.

There are critics who feel that intelligence tests are inappropriate for American Indians as well. Havinghurst (1970) concluded that Indians have a different value system and tend to interact in noncompetitive ways. Trying to excel on a given test may not be a meaningful goal for them. Heath and Nielson (1974) point out that Indians tend to use non-verbal communication techniques, and

that their language structure differs quite dramatically from English. They feel that tests requiring advanced verbal ability would be inappropriate for natives. Cress (1974) points out that Indians tend to respect their elders, and to adopt their priorities. He suggests that achievement in white schools has not usually been highly valued.

Alternate Minority Viewpoint:

There appears to be no great outcry against I.Q. testing of children from French Canadian or European backgrounds. One explanation for this, cound be that the creators of the tests have backgrounds similar to these students, and therefore, these students are not faced with an assessment from a different culture. Another possible reason is that they do quite well on both achievement and intelligence tests. Peal and Lambert (1962), in their study of matched bilinguals (French), and unilinguals (English), found that the bilinguals outperformed the unilinguals on both verbal and performance tests. They conclude that this may be due to the enriched cultural experiences of bilingual students. They also suggest that children who have experiences in more than one language develop their abstract reasoning ability more quickly, and that they acquire greater flexibility in their reasoning processes. Liedke and Neilson (1968) matched grade one students on a Piagetian concept formation task, and concluded that the bilingual children have mental processes more advanced than unilinguals.

Lambert and Tucker (1973) reported on a longitudinal study of children from English homes attended French kindergarten and grade school. They found that the students' performances were significantly above those attained by students who went to English schools. Cummins (1974) matched students for social status, age and sex. He found

bilinguals to have better verbal and general reasoning skills.

In his research, Kittel (1963) did not use a measure that would establish the level of bilingualism of a child. Instead, he established whether the parents spoke another language, and classified the children of parents who did, as coming from a bilingual environment. He assessed bilingual students in Grade three and again when they were in Grade five. He concluded that at the Grade three level, assessment of bilinguals should be approached with caution, as they might not yet have adjusted to the switching back and forth between cultures. By Grade five, Kittel felt test results would generally indicate that the children from a bilingual background were performing at least as well as the unilinguals.

Ewanyshyn (1978) analyzed the effects of a Ukranian emmersion program, on students who were predominantly English speaking, and concluded that no detrimental effects occurred in the areas of intellectual development, or academic achievements. In fact, the students' achievement was considered satisfactory in both their English skills and their Ikranian language ability. Moss (1979) studied the effects of learning another language (Hebrew), on the reasoning abilities of the students. She referred to these students as "pseudo bilinguals" as they were just in the initial stages of learning Hebrew. Moss concludes that even with only part time experience in another language, students tend to show superior reasoning ability over those who are unilingual. The results of present studies seem to indicate that bilinguals, and those who come from bilingual environmental backgrounds, benefit in their intellectual development from the experience.

Suggested Alternatives to Present Intelligence Tests:

The criticisms of the present intelligence tests, that suggest that certain individuals, or races are born with more limited capabilities, are quite understandable. As Kamin (1975) points out, no one would wish to be categorized as intellectually inferior, as this carries a social stigma. However, as the NAACP Report on Minority Testing (1976) indicates, some form of assessment is necessary. The critics, who claim that the present tests are biased against minority group members, have suggested alternatives. Some of the suggestions for improving the present testing methods amount to modifying the existing instruments and procedures. Hynd and Garcia (1979), in a study of Navajo Indians using the WISC-R, obtained similar results as did Cundick (1970) studying southwest Indians, and St. John et. al. (1976) studying Indians in northern Ontario. The Indians in all three studies performed better on the performance, than on the verbal tests. They suggested that in the assessment of Indians, the two scales not be combined. They also recommend that the verbal score be used to judge and measure school achievement, and the performance score be viewed as the individual's intellectual potential. In their study of Indians in British Columbia, Seyfort et. al. (1980) recommends extreme caution in the interpretation of subtest scores. According to their analysis of the data, a number of items do not contribute significantly to the total test variance. They feel that there is a danger of overinterpretation of test results. B.S. Pray (1979) proposed that four subtests of the WISC-R that are the most heavily culturally laden, Information, Comprehension, Vocabulary and Picture Arrangement, need modification in their scoring system if they are to be useful in assessing handicapped Indian children. He feels the formula he has

devised would be useful in this area. Hays and Smith (1980) recommend that a Raven's be used in conjunction with a WISC-R, when assessing juvenile delinquents who are members of a minority group. They feel each test measures different intelligence factors, and by combining the two, a fairer estimation of an individual's intelligence can be obtained.

Other researchers have gone further with their recommendations. In their studies of the Indians and Metis in Alberta, MacArthur (1962) and West (1962) both conclude that intelligence tests with a verbal component are not appropriate for the population, as they are biased against the Indians and Metis. They suggest that a nonverbal test, such as the Raven's Standard Progressive Matrices, with its heavy "g" factor loading, is a more appropriate instrument. MacDonald and Netherton (1969) who performed studies in approximately the same region, reached similar conclusions.

In their variation of testing procedure, Carlson and Wiedl (1980) incorporate a learning component. They call this procedure "dynamic feedback" and, as it combines learning with assessment, they feel the procedure offers compensatory gains for minority children. Williams (1971), dissatisfied with the available assessment instruments created the Black Intelligence Test for Cultural Homogeneity (BITCH). It consists of vocabulary and expressions common to ghetto children. The author claims it is an intelligence test for poor blacks and a sensitivity instrument for whites (BITCH, in Buros, 1975). deAvila (1974) based his assessment procedure on the stages of intellectual development that had been proposed by Piaget. He stated that, as every child went through these stages, his assessment procedure would be less

biased than presently available tests. Calling the procedure the Program Assessment of Pupil Instruction (PAPI), he devised four measures, a cartoon conservation scale, a water level task, a figural interactions test and a social test. It is yet to be determined how adequately this procedure will surplant present tests.

Mercer (1977) has developed an extensive assessment procedure which includes measures of an individual's health, socio-economic status (S.E.S.), family environment and a WISC-R score. The procedure is called the System of Multicultural Pluralistic Assessment (SOMPA). She is of the opinion that the only reasonable way to arrive an an estimation of a person's potential mental abilities is by considering all factors that can influence their development. The SOMPA procedure, according to Greenleaf and Smith (1978), has rectified the cultural bias previously associated with special class placements in Louisiana. Though SOMPA was developed with a Spanish speaking population as the reference group, More and Olderidge (1980) feel that the procedure holds promise for use with Indian children as well.

Defence of Present Intelligence Tests:

Though psychologists claiming cultural bias have been vocal, others who disagree have also come forward. Jensen (1975) presented a statistical analysis of a series of tests administered to blacks and whites. He found no difference between the groups in the internal consistency of their responses, reliability, error distractor choice, or item difficulty order. He concluded that the WISC-R, Stanford-Binet and Raven's Progressive Matrices Test are not biased against blacks. In his research, Miele (1979) found no evidence of bias in his factor structure analysis of black and white performance on the WISC-R.

He concluded that the observed differences were due to the maturity level defferences rather than test bias. In their research with black and white retarded children, Richmond and Long (1977) concluded that there was no evidence of cultural bias in the WISC-R Verbal, Performance, or Full Scale scores.

Factor analysis and analysis of variance have led others to similar conclusions. Dean (1979) pre and post tested whites, and a group of Mexican-American children, using the WISC-R. He concluded that the WISC-R was reliable and met predictive validity requirements for both groups. In a more recent study of Mexican-American children, Dean (1980) came to the same conclusion of no evidence of bias in the WISC-R.

Reschly (1978) examined the factor structure of WISC-R test results for four groups: Anglos, Blacks, Chicanos and Native-American Papagoans. According to Reschly, the scales appeared appropriate for all the groups and these results added to his confidence in the construct validity of the WISC-R.

Other WISC-R studies have also concluded that it is an unbiased test. Sandoval (1979), in a study of Anglos, Blacks and Mexican-Americans, Reschly and Reschly (1979) studying Anglos, Blacks, Chicanos, and Native Papagoans, and Gutkin and Reynold (1980), in their study of Anglos and Chicanos referred for psychological services, agree on the utility and lack of bias of the WISC-R.

Criticisms of Suggested Alternatives:

The suggested alternatives to the present intelligence tests have also been criticized. The BITCH was used in one study by Long and Anthony (1974). Though the instrument was supposed to be fairer to

blacks, the researchers found that it didn't differentiate EMR candidates any differently than the WISC-R.

The SOMPA has been criticized on both a research and conceptual level. It is suggested by Salvia and Ysseldyke (1981) that after close scrutiny of the SOMPA, it is at best an experimental research instrument. Oakland (1980) attempted to predict achievement scores for two S.E.S. levels, low and average, with four social groups, Anglos, Blacks, Mexican-Americans and Chicanos. He found that the predictive validity decreased with the inclusion of the SOMPA measures. On a more fundamental level, Goodman (1979) considers that an instrument that alleges to measure potential is a regressive step in assessment development. In her opinion, psychologists have been making steady progress in developing instruments that measure mental ability more and more accurately. To return to archaic assumptions that somehow potential can be measured, is counterproductive, according to Goodman.

Crawford (1979) perhaps offers the most satisfactory summation with regard to cultured bias-controversy. He is of the opinion that both the environmentalists and geneticists have taken positions that are too extreme. He feels that between these extremes, there is room for consensus on the issue.

Uses and Consequences of Testing

Test uses and test outcomes result in actions, and these actions are accompanied by consequences. As Messick (1980) points out, it is not sufficient to establish the reliability and validity of a test statistically, as well, the use of the measures and the consequences of that use must be considered. In the previous sections, the question of the appropriateness of the tests was explored. In this section,

the purposes for testing and the results of testing will be examined.

An asserted purpose for intelligence testing is to allow more insight into the relative strengths and weaknesses of an individual so that his educational and career opportunities can be maximized (Page, 1980). Salvia and Ysseldyke (1981) feel that a major use of test results is as a screening instrument for potential deviance. According to them, a test can be used to identify students who are not, or will not benefit from their present placement, or would benefit more from another placement. Examples of these uses are Winger (1968), and Chismar (1971), who used the Lorge-Thorndike to identify underachievers in their separate studies.

In studies in which the focus is the identification of special students, varying degrees of success have been achieved using different assessment methods. Ellis (1969), in comparing the groups the Lorge-Thorndike and WISC-R identified as underachievers, though dealing with only white meddle class children in the sixth grade, found that the two instruments did not totally identify the same population. In a study by Keech (1966) gifted students were identified with the Stanford-Binet. The teachers of these students were only able to identify 58% of those that the Stanford-Binet had indicated were gifted. In a related study, Kundel (1966) identified low I.Q. children with the Stanford-Binet and found that teachers were only able to identify 48% of these. In a study by Skager and Fitz-Gibbon (1972), Raven's Standard Progressive Matrices was administered to a group of students, followed by Raven's Advanced Progressive Matrices. In the group of fourteen students identified as possible gifted by these two tests, the WISC-R identified eight as gifted. Teachers who were asked to identify the

gifted in the same group identified four. Thus, it would appear that relying on teacher's observations alone could result in the nondetection of special students.

A major concern with those who oppose present assessment procedures is that racist sentiments may arise out of the research. Kamin (1975) points out that one of the results of irresponsible testing and drawing of conclusions in the 1920's, was that individuals in the federal government attempted to restrict the immigration of specific nationalities and races. He suggests that if the various legislative bodies take the position that intelligence is inherited, and minimally influenced by environment, they may neglect or even cancel programs intended to help children in deprived environments. The fear that programs such as Head Start may be cancelled, is shared by Cronbach (1975) and Conwell (1980).

A more direct criticism of testing and its use, comes from Cardinal (1969), who claims that it is an attempt of educators to make a second class white out of the Canadian Indian. He feels that the process is a constant undermining of the Indian culture and values by educators and, simultaneously, and attempt to instill white standards in their place.

Adelberto (1970) in the United States, criticizes assessments from a different perspective. He claims that the government, rather than attempting to assimilate the Spanish speaking people, is trying to keep them in their present place. He feels that this maintaining of the status quo has proven detrimental to the advancement of his people. A similar point on behalf of the blacks in America had earlier been made by Williams (1970).

There are also more specific criticisms of the consequences of present day assessment techniques. Mercer (1974) points out that due to testing for placement in special classes, there is an over-representation of Mexican-Americans, blacks and other minority group members in special classes. She feels that these special class placements result in lower teachers' expectations of the students, leading to an inferior education of the students, and resulting in limiting their higher educational and career opportunities for life. Mercer, and others who felt the same way, were so cogent in their arguments, that as a result, the courts have terminated the intellectual assessment of minority group members in the San Francisco schools (Cohen, 1977).

At the classroom level, critics of intelligence testing are concerned with the evident importance teachers place on test results (Matuzzek and Oakland, 1979), (Smith and Knoff, 1981). In a recent study Cuttance (1980) found that teachers placed a great deal of importance on the previous years' results, found in the cummulative records of the students. The teachers felt that these results were valuable in setting educational objectives for their pupils. Cuttance feels these findings indicate the importance of having only accurate, reliable and appropriate assessment instruments used, as the assessment results will be recorded and future teachers will rely on that information.

There is criticism of testing at the individual level as well. The consequences of being labelled are well documented by individuals such as Hobbs (1975), Fine (1975), Fuller (1977), and Mercer (1973a). They agree that labelling someone as being mentally retarded, or

handicapped is a stigma that they may carry the rest of their lives. They feel it affects a person's feeling of self worth, the level of his aspirations, the way he relates to others, and the quality of his life in general. Tucker (1980) points out that as a result of these criticisms of over-representation of minorities in special classes, the trend is towards classifying the same student's "Learning Disabled". He feels that a relabelling process has begun, with the intention of removing large proportions of the minority group students out of the regular system.

In defending the intelligence testing of minority group members, Green (1978) points out that the testing is useful as it reflects the shortcomings of the present system in meeting minority group needs. Jensen (1975a) defends the testing on the basis that the results should lead to the development, and implementation of appropriate educational objectives. Sattler seems to sum up this position well when he states"

> "Test scores should not be accepted as fixed levels of either performance or potential; instead, they may be used to determine the magnitude of the deprivation that is to be overcome by a planned program of remedial activities. Scores can also be used to compare disadvantaged children with one another. Still another way in which scores can be useful is to compare the child's current test performance with his previous test performance. In the last analysis the examiner and other test users must accept the responsibilities involved in interpreting and in using educational and psychological tests". Sattler, 1974, p.46.

Implications of the Review of the Literature for this Study

Proposed theories of intelligence appear to be coming closer in their positions. The consensus appears to be that there is a general intelligence factor in each intelligence test which is present to a greater or lesser degree, depending on the nature of the test. The second point of agreement found in intelligence theories is that both the environment

and heredity contribute to the development of intelligence, but to what degree is still an unresolved issue.

A premise that seems to be generally accepted is that intelligence tests do not measure innate potential. There is also agreement that intelligence tests measure specific mental abilities, and not intelligence per se. The mental abilities measured are not fixed in time, and are subject to environmental influences. Group intelligence tests are quick and convenient ways of arriving at an approximate measure of intelligence or for group comparison purposes, but they lack the clinical accuracy of individual assessments. The importance of an I.Q. score should not be overestimated and should be viewed in context with other behavioral and ability measures.

A social issue of primary importance in intelligence testing is the question of the appropriateness of the test for minority groups. Though the controversy is by no means resolved, it appears that certain culture-reduced tests and the WISC-R and Stanford-Binet are reliable instruments. However, it appears necessary to use extreme caution in scoring and interpreting the results when the tests are used with most minority groups. An exception occurs with groups who have a European background or French Canadian background. Though caution should be used when assessing the intelligence of young bilinguals, they do not appear to be handicapped on their test results by being exposed to another culture. In fact, their later scores tend to be higher than those of unilinguals.

> "The validity of a measurement consists in what it is able to accomplish, or more accurately, in what we are able to do with it. The basic question is always whether the measures have been so arrived at that they can serve effectively as means to the given end." (Kaplan, 1964, p.198)

The fear of critics of intelligence testing is that the end result will be a retention of the status quo for minority groups, allowing limited opportunities for upward mobility. They are concerned about races being labelled intellectually inferior and about minority group children being treated, and educated, as inferiors. They strongly oppose children being labelled as mentally inferior, especially when the children come to accept the labels themselves.

Those involved in testing claim their motivation for testing is to better understand the individual and thus, be more responsive to their needs. There are two questions not yet resolved. What specifically are the intents of the individuals who endorse intelligence testing and use the test results, and are concerns of the critics of intelligence testing justified?

Specific Significance of the Review for this Study:

Implications of the review of the literature suggest that all intelligence tests contain a general intelligence factor. Further, each intelligence test measures specific mental abilities. Thus, when different intelligence tests are used as screening devices, generally nonidentical populations are identified.

A review of the literature also indicated that minority groups tend to score lower on standard intelligence measures. This could be due to genetic influence or environmental factors. The consensus is that it is likely an interaction of both. Native children in particular perform more poorly on tests requiring verbal abilities, than on nonverbal

intelligence tests. Culture fair tests such as Raven's Standard Progressive Matrices appear to be less biased in assessing the general intelligence of native students.

The minority groups that are the exception, that is, that do not appear to be penalized by intelligence testing, are those of French Canadian or European language backgrounds.

The importance of careful consideration before administering any intelligence test, especially when members of a minority group are to be assessed, has been emphasized in the literature. It is evident that before a test is used it should at least meet the following criteria:

The test should back up its claim, of testing specific abilities, with statistically significant information, concerning its validity and reliability, on the basis of studies done using the test. The test must be appropriate in terms of age, maturity level, and administrative format, for the group to be assessed. The results of the tests have to be useful, that is, they must be in a form that allows them to be combined with other information, in planning teaching strategies, and setting educational goals.

CHAPTER 3

The summary and implications of Chapter 2 underline the importance of a critical examination of an intelligence test before it is administered and the results recorded. The Lorge-Thorndike will be examined in the light of how adequately it identifies underachievers (UAs), in a population of students from different cultural and language backgrounds. The primary question is concern regarding many beta type errors occur when the L-T is used as a screening instrument. Raven's Standard Progressive Matrices will also be assessed for its utility in identifying UAs. This chapter will review the data gathering procedure, describe the population, review the measurement instruments, and describe the data analysis procedure. In Table 1, located at the end of this chapter, will be found a list of abbreviations, accompanied by their definitions and references as used in this study.

Population

The population consisted of students in Grades 3 through 8, in attendance at three Catholic schools in the Cold Lake - Grand Centre area. These three schools form the eastern extremity of Lakeland Catholic School Division No. 150.

The study was comprised of 341 students, of whom 332 had completed data. Of the remaining nine, six moved and three had incomplete data. Of the 332 students, 61 were in Grade 3, 55 were in Grade 4, 55 were in Grade 5, 55 were in Grade 6, 54 were in Grade 7 and 52 were in Grade 8.

Data Gathering Procedure

An informal survey revealed that fewer than 20% of the students were conversant in another language. Rather than attempting to determine the degree of biliguality of each student, Kittel's (1963) concept of

bilingual background was used. Kittel's premise was that children are influenced directly and indirectly by their parents' knowledge of more than one language. In agreement with this premise, the parents of the students were contacted and queried as to the extent of their knowledge of another language. The two criteria identifying a student of bilingual background were: Whether the parents knew another language, and whether they spoke it at home.

On the basis of parental reports, the students were classified into five groups; Group 1 - the English speaking unilinguals, Group 2 - those from a French bilingual background, Group 3 - those with a Cree language background, Group 4 - the students whose parents spoke Chipweyan, and Group 5 - those whose parents spoke a language other than the one specifically mentioned. Of the 332 students, 111 were classified Group 1, 67 classified as Group 2, 86 classified as Group 3, 31 classified as Group 4 and 37 classified as Group 5.

Annually, in May, the students in these three schools, in Grades 3 through 8, are assessed on the Canadian Tests of Basic Skills (C.T.B.S.). With the consent of Lakeland Catholic School Division, this year the same students were also assessed on the Lorge-Thorndike and the Raven's. As the testing involved seven sittings per grade, care was taken to avoid assessing elementary students more than once per day. In Grades 7 and 8 the maximum of a morning and an afternoon test occurred one time only. The students were well supervised with monitors constantly making certain that the students observed proper procedure, and were not making clerical errors.

The instructions were read aloud and in a consistent manner, as the administrator was very familiar with the instruments. The test settings were comfortable and, given the fact that these were group tests, relatively free of distractions. In this study, the Wechsler Intelligence Scale for Children - Revised (WISC-R) is used as the criterion reference to compare the adequacies of the Raven's and Lorge-Thorndike in identifying UAs. An individual is assessed using the WISC-R for one or more than one of three reasons: 1) when a student's progress is erratic or inconsistent with a teacher's expectations, a referral is submitted requesting an individual assessment, 2) students are assessed as a screening device to ascertain their appropriate intelligence category. When a student's performance is weak on both teacher made and standardized tests, he is assessed to see if this may be due to a specific weakness, or to an overall weakness of his mental abilities, and 3) students who have been placed in a remediation setting or a resource room are regularly reassessed, to observe whether any significant changes have occurred in their overall intellectual ability or in their pattern of strengths and weaknesses.

Of the 332 students, 77 have been assessed with the WISC-R; 27 of the 111 in Group 1 (24%), 16 of the 67 in Group 2 (24%), 16 of the 86 in Group 3 (19%), 11 of the 31 in Group 4 (35%), and seven of the 37 in Group 5 (16%).

The Measurement Instruments

The C.T.B.S. is used in this study as the standardized, norm referenced achievement measure, to which an individual's intelligence test performance can be compared. The WISC-R is the criterion reference, by which the other two intelligence measures, the Lorge-Thorndike and

the Raven's, will be judged in terms of adequacy in identifying UAs.

The Canadian Tests of Basic Skills (C.T.B.S.)

The C.T.B.S. is a Canadian achievement test based on the well established Iowa Test of Basic Skills which was developed in 1935. It was modified to reflect Canadian content and objectives in its subject areas. It was normed on representative sample of the Canadian population (Hieronymus and King, 1975).

The C.T.B.S. is composed of tests in the areas of; Vocabulary, Reading Comprehension, Language (four subtests), Work Study Skills (three subtests) and Mathematics (two subtests). The editors of the tests, Hieronymus and King, state that the tests are not intended as measures of subject content. Their intent is to measure only generalized intellectual skills and abilities. Their reasoning for not supplying subject specific test is:"The great heterogeneity of school-to-school variability, in curriculum organization, and content also makes it impossible to supply tests in these special subjects that are well adapted to most local situations." (p.6)

According to the editors, the reliability of the C.T.B.S. was a major consideration in the construction of their tests. To ensure reliability, they have made the battery longer than most achievement test batteries. Their split-half reliability coefficients are based on populations ranging from 406 to 540 at each grade level. The reliability coefficients range between .86 and .89 on the Vocabulary test, between .91 and .93 on the Reading Comprehension test, between .94 and .96 on the total Language test, between .90 and .93 on the total Work Study Skills test and between .88 and .91 on the total Mathematics test (Hieronymus and King, 1975, pp. 52-54).

The editors have assumed Cronbach's position in their discussion of the validity of the C.T.B.S. Cronbach had stated that "validity is the task of the test interpreter" (in Hieronymus and King, 1975, p.40). The editors feel that in order to ensure the suitability of the test for any given region or school, the perspective user should take the testl. That is the items and tests should be judged as to their appropriateness for the students being assessed. The editors have added, that before a measure to which it is compared should be noticeably superior. As they have not presented comparative validation statements, it appears that they have not found a superior battery with which to compare theirs. This may be due in part to the fact that there is no comprehensive Canadian achievement battery, other than the C.T.B.S., at the present time. The teachers in the three schools involved, following the advice of the editors of the C.T.B.S., took the test. They reached a consensus that the Work Study Skills test was not appropriate in both content and difficulty level. It has, therefore, been omitted from the test battery. The teachers further concluded that the battery was both appropriate and well constructed.

The Wechsler Intelligence Scale for Children - Revised (WISC-R)

The WISC-R consists of 12 subtests, six in the Verbal Battery and six in the Performance Battery. The Mazes in the Performance Battery and the Digit Span in the Verbal Battery are not used in arriving at the Deviation I.Q. (D.I.Q.). The Digit Span test is quite often used as a supplemental measure of an individual's short term auditory memory ability. The five subtests that compose the Verbal Scale and are used in arriving at the D.I.Q. are; Information, Comprehension, Arithmetic, Similarities and Vocabulary. The five subtests that compose the

Performance Scale are; Picture Completion, Picture Arrangement, Block Design, Object Assembly and Coding.

The raw scores on each subtest are converted into scaled scores that range from values of one to 20. The total of the five scaled scores in the Verbal Battery are converted into a Verbal scale D.I.Q. with a mean of 100 and standard deviation of 15. The same process is repeated to arrive at the Performance scale D.I.Q. The ten subtest scaled scores are totalled and converted into the Full Scale D.I.Q. (Wechsler, 1974).

In creating the WISC-R, Wechsler intended his test to measure the overall capacity of an individual. The test was to assess an individual's ability to understand, and cope with the world around him. He did not establish an hierarchical order for the subtests, as he felt they all were necessary components of the test (Kehn, 1975).

The reliability of the test was based on the ten subtests used in establishing the D.I.Q.s. Split-half reliability coefficients were reported for the 71/2, 101/2 and 131/2 age levels. The coefficients ranges are; from .92 to .95 for the Full Scale, .88 to .96 for the Verbal scale and .86 to .90 for the Performance scale.

Sattler (1974), Jensen (1975), and Dean (1980) review numerous studies attesting to the reliability of the WISC-R and its valid use as an assessment instrument with numbers of minority groups. The major possible drawbacks of the WISC-R include the possibility of assessor bias (Mercer, 1972), the subtest results may be overinterpreted (Seyfort et.al., 1980), and the focus of attention may be on the Full Scale score, without due attention being given to the separate Verbal and Performance scores (Hynd, 1979).

The Canadian Lorge-Thorndike Intelligence Tests (L-T)

The L-T is intended as a measure of abstract intelligence which is defined by the authors as "the ability to work with ideas and the relationships among ideas" (Lorge et.al., 1967, p.3). The authors feel that abstract intelligence is closely linked to academic achievement.

The Canadian version of the L-T was adapted from the American version by E.N. Wright. It was standardized and normed on the same Canadian population as the C.T.B.S. for grades 3 through 8 as part of an integrated program.

The L-T consists of a Verbal and Nonverbal Battery. The Verbal Battery contains five subtests: Vocabulary, Verbal Classification, Sentence Completion, Arithmetic Reasoning and Verbal Analogy. Each subtest is seven minutes in duration. The Nonverbal Battery does not rely on reading abilities, as it is composed of pictoral and numerical items. It has three subtests: Picture Classification, Pictoral Analogy and Numerical Relationship. Each of these subtests are nine minutes long.

There are six levels of this test ranging from A to F. The levels are generally meant to correspond with grades starting with the use of level A in Grade 3. The authors suggest that a given area may choose to alter this procedure, when they take the socio-economic status of the community, and the ability level of the students into consideration. For example, the authors recommend using level A with a Grade 4 or 5 class if the socio-economic status of the community is low and the students demonstrate below average academic ability.

The manual supplies tables for each test level and battery to convert raw scores to D.I.Q.s. The L-T Full Scale I.Q. is computed by

adding the Verbal and Nonverbal I.Q.s, then dividing by 2.

The authors provide odd-even reliability statistics for each level of both batteries. They are based on populations ranging from 511 students in Grade 3 (level A) to 598 students in Grade 7 (level E). The reliabilities for the Verbal Battery range from a high of .95 at the Grade 3-A level, to a low of .83 at the Grade 8-F level. The Nonverbal reliabilities range from .93 at the Grade 3-A level, to .89 at the Grade 9-F level. The authors state that as the intercorrelations between the Batteries are lower than the reliabilities of the Batteries (ranging from .68 to .55), results of the Verbal and Nonverbal Batteries, as well as the Full Scale I.Q. score (Lorge et. al. 1967, p.29). The authors of the manual offer no validity data using a Canadian population. They base the validity of the L-T on its correlations with other older tests in America. They state that the L-T Verbal Battery correlates in the high 70's and low 80's with the WISC-R Verbal Scale, the Stanford-Binet and the Verbal Reasoning and Numerical Abilities sections of the Differential Aptitude Tests. They also state that the L-T Nonverbal Battery correlates in the high 60's and low 70's with the same tests (Lorge et. al., 1967, p.29).

The L-T has been used in numerous studies. West and MacArthur (1964) found the Nonverbal Battery appropriate for the intellectual assessment of Indian and Metis. Purl and Curtic (1970) concluded on the basis of their study that the L-T was a better predictor of academic achievement with minorities than either the Raven's or WISC-R. Fisk (1979) used the L-T as the intelligence criterion for identifying the learning disabled. From the literature, it would appear that the L-T is a well established intellectual assessment instrument. Its appropriateness, and/or usefulness, in reference to the population in

this study has yet to be determined.

Raven's Standard Progressive Matrices (the Raven's)

The Raven's consists of 60 problems divided into five sets of 12. The first problem in each set is quite simple. As the problems in each set increase in difficulty, they offer a learning experience to the test take, indicating to him the operations required to answer successive problems. Each problem consists of a large, boldly presented pattern with one peice missing. The individual chooses from one of the alternatives presented on the bottom and records the number of his choice on a separate answer sheet. There is a minimal amount of instruction needed in group testing situations as the first problem is worked out very carefully and thoroughly with the whole group. Scoring is simple as it consists of totalling up the number of correct responses. There is not time limit for the test, but everyone is usually finished in less than an hour.

The author describes the Raven's as a "test of observation and clear thinking." (Raven,1960, p.2). He feels that it should be used in conjunction with the Mill Hill Vocabulary Scale in the assessment of an individual. According to the author, it is not in itself a test of general intelligence. He does, however, mentioned in the same paragraph that it was found to have a "g" loading of .82. Raven states the test-retest reliability of the instrument varies with agr group from .83 to .93. He validated the Raven's by correlating it with the Terman-Merill scale and found correlation to be .86. Burtner (in Buros,1975), in his review of the Raven's, found it to be a useful assessment instrument. He felt it would be particularly appropriate for assessing members of ethnic or minority groups, and individuals who have difficulty

communicating. MacArthur (1967), in his study of Eskimos and Metis, used the Raven's without the Mill Hill Vocabulary test. He found the Raven's a more culture fair way of assessing the intelligence of these two groups. Jensen (1970), in his review of the studies where the Raven's had been used, came to the conclusion that it was a reliable and unbiased measure of "g". He felt that it was valid to use the Raven's with any minority group. As the literature indicated that the Raven's was a widely used test with minority groups, it was included in this study to answer two basic questions: Does the Raven's identify the same students as underachievers as the L-T? Does the Raven's identify the same students as UAs as the WISC-R ?

Data Analysis Procedure

The students from the three schools were initially grouped by grade. Their raw scores on the C.T.B.S. and Raven's were normed on their grade group and converted into z scores. Z scores are standard scores with a mena of 0.0 and a standard deviation of 1.0. The L-T Verbal and Nonverbal (L-T (V) and L-T (NV), respectively) I.Q. scores were also normed for each grade and converted into z scores. The only scores not normed by grade were the WISC-R D.I.Q.s. The decision to retain the mean at 100 and the standard deviation at 15 was made for two reasons: The number for each grade was extremely small. Further, the students assessed on the WISC-R were not a random sample, and were atypical due to the selection process.

The standard scores for each individual having been established, the students were regrouped in accordance with the five language classifications. Pearson product-moment correlations were performed to determine the correlation coefficients (r), that is, the amount of

agreement between the various intelligence tests and the C.T.B.S. results. The means and standard deviations of each group in z scores were established. A one way analysis of variance (ANOVA) was done using the Scheffe procedure for uneven groups. This was to determine whether there was a significant difference in means between groups on the C.T.B.S., Raven's, L-T (V) and L-T (NV).

The next step was to subtract each individual's score on the C.T.B.S. from their score on the Raven's. If the difference was -1.0 or more negative, the individual was classified as an OA (Overachiever) as his achievement score significantly exceeded his intelligence result. If the difference was 1.0 or greater, he was classified as an UA (Underachiever), as his intelligence score significantly exceeded his achievement result. Those whose scores fell between the two extremes were classified as NSDs (Their results revealed No Significant Difference between intelligence and achievement). This same procedure was fallowed for; the L-T (V) - C.T.B.S., L-T (NV) - C.T.B.S., L-T Full Scale (FS) - C.T.B.S., WISC-R Verbal (V) - C.T.B.S., WISC-R Performance (P) - C.T.B.S., and WISC-R (FS) - C.T.B.S.

Using the categories of OA, NSD and UA, a series of cross tabulations was performed to determine the extent of agreement of the intelligence measures classifications. The final procedure was the creation of scattergrams. They depict the degree of agreement of the WISC-R (FS), L-T (FS), and the Raven's more graphically.

TABLE 1

Table of Abbreviations and Symbols Used in This Study.

ANOVA	=	Analysis of variance (one way and using Scheffe procedure for uneven groups)
C.C.A.T.	=	Canadian Cognitive Abilities Tests (a fairly recent group intelligence test normed in Canada)
C.T.B.S.	=	Canadian Test of Basic Skills (a norm referenced Canadian achievement battery)
D.I.Q.	=	Deviation Intelligence Quotient
(F.S.)	=	Full Scale (Used with both the WISC-R and Lorge-Thorndike)
L-T	=	The Canadian Lorge-Thorndike Intelligence Tests (group tests)
N	=	Number
N.S.D.	=	No Significant Difference (less than one Standard Score between achievement and intelligence test results)
(NV)	=	Non-verbal Battery (with the L-T)
OA	=	Overachiever (a student whose standard score on an achievement test exceeds his standard score on an intelligence test by 1 standard deviation or more)
(P)	=	Performance Scale (with the WISC-R)
r	=	Pearson Product - moment correlation coefficient
The Raven's	=	The Standard Progressive Matrices (author J. C. Raven)
Sig.	=	Significance
Stanford-Binet	=	Stanford-Binet Intelligence Test (individually administered)
UA	=	Standard Score on intelligence exceeds standard score on achievement
(V)	=	Verbal Scale (when used in conjunction with the WISC-R Verbal Battery (when used in conjunction with the L-T)
WISC-R	=	Wechsler Intelligence Scale for Children-Revised (individually administered test)

CHAPTER 4

Analysis of the Results

This chapter summarizes the results of the data analysis performed in this study. The initial data to be summarized is presented in Table 2 where the means and standard deviations of the C.T.B.S. and Raven's raw scores, and L-T (V), and (NV) D.I.Q. scores are presented for each grade. This summary is followed by one focussing on the standard score means, and standard deviations of the five language groups. The summary will also include the ANOVA of the group means on the different tests. A section will follow examining the extent of the correlations of the measures for the language groups. The final section will summarize the crosstabulations that result when different intelligence measures are used in arriving at the OA, NSD, and UA classifications of students.

Means and Standard Deviations of Grouped Data by Grade

The data in Table 2 summarizes the scores on each variable by grade. Three trends are evident in this summary. The first trend appears to indicate that the C.T.B.S. mean scores improve by grade, as they increase from a mean of 71.39 at the grade 3 level, to a mean of 98.24 at the grade 7 level. The increase in the total number of items at each level accounts for most of the difference in raw score means. The percentages in brackets are included to indicate the mean percentage of correct responses at each grade level. The data indicates that the greatest difference between percentage means, at the extremes, is less than 2 percentage points, or less than 8 percent. Appendix A includes distribution charts for the various assessment instruments.

TABLE 2

Means and Standard Deviations of Achievement and I.Q. Raw Scores By Grades

	Grade	Mean	Standard Deviation	% of Item Correct
C.T.B.S.				
	3 N = 61	71.39	28.76	25%
	4 N = 55	81.04	31.96	26%
	5 N = 55	92.29	31.37	27%
	6 N = 55	97.71	31.42	27%
	7 N = 54	98.24	32.25	26%
	8 N = 52	95.69	30.94	25%
RAVEN'S				
	3	29.90	10.23	
	4	35.56	10.54	
	5	39.58	9.48	
	6	40.50	9.05	
	7	42.80	7.57	
	8	44.67	7.97	
L - T (V)				
	3	93.26	19.93	
	4	97.28	18.38	
	5	94.09	15.03	
	6	93.25	17.76	
	7	94.87	15.45	
	8	92.81	14.53	
L - T (NV)				
	3	96.62	19.61	
	4	104.02	21.64	
	5	101.38	15.78	
	6	102.86	18.66	
	7	110.93	18.05	
	8	104.35	14.40	

A second trend that is apparent from the data, is that on the Raven's the mean number of correct responses increases with each grade level. This difference is not attributable to a difference in item totals, as every grade takes the identical test. An additional feature of this trend is that the mean difference from one grade level to the next decreases with each subsequent grade. The difference between the means for grades 3 and 4 is approximately 6, while the difference between grades 7 and 8 is less than 2. The simultaneous trends of increased accuracy and decreasing difference with each higher age level, are consistent with the test norms, as presented in the manual (Raven, 1960). The norms however, do not appear to be appropriate for the population in this study, especially in the lower grades. The norms indicate that a 9 year old attaining a score of 24 would be at the 50th percentile. Though means are not directly interpretable as percentiles, for the grade 3 students, whose mean age is slightly less than 9 years, the mean score was slightly less than 30. This difference between the published norms and the results of this study tend to support MacArthur's (1962) contention that locally developed norms are needed for the Raven's.

Another notable trend that is evident in the data in Table 2 is that at each grade level, the students perform better on the L-T (NV), than they did on the L-T (V). Hynd (1979), among others, had noted that this trend has existed for Indian children, for the data from this study suggests that superior (NV) to (V) scores occur generally throughout this population of students. An additional observation is that the difference between the (NV), and (V) tends to increase with each grade level. At the grade 3 level, the difference is approximately 3 points while at the grade 8 level,the difference is approximately 12

points. This report approximates that reported by Sattler (1974) in his research.

In concluding this review of the graded data, two additional points are deserving of mention. All of the D.I.Q. means are less than 11 points from the standard mean of 100, and the mean percentage of correct responses does not significantly differ between any two grades. This indicates that academic and intellectual ability is fairly evenly distributed throughout the grades.

Standard Score Analysis

For the analysis that follows, the students were reorganized into their language background categories. The data in Table 3 indicates a striking contrast between group performances on various measures. The z score means on the various tests reveal that group 3, comprised primarily of the mostly Metis Cree, and group 4, primarily Indian Chipewyan, are the only groups who have negative means on every measure. In fact, except for group 1's performance on the Raven's, groups 3 and 4 were the only groups to have negative means on any of the measures. The ANOVA data accentuates the fact that group 3 had a significantly lower mean than at least one other group, on each of the tests. Their means on the C.T.B.S., L-T (V) and L-T (NV) were significantly lower than the means of groups 1, 2, and 5. The only group that didn't have means with significant differences from group 3's, was group 4. Group 4 also had means lower than the means of group 1, 2, and 5. The mean difference did not reach the .05 level of significance when they were compared to groups 1, 2, and 5 on the Raven's, nor when they were compared to group 1 on the L-T (NV). The data in Table 3 makes it very evident that the native children from Cree and Chipweyan environments

TABLE 3

Standard Score Means, Standard Deviations and ANOVA For All Group Tests

TEST TITLE	GROUP	z Score $\bar{X}$	S.D.	GROUP 1	2	3	4	5	ANOVA * F Probability
CTBS	1								
N = 111	English	0.158	1.064			*	*		0.0000
	2								
N = 67	French	0.375	0.821			*	*		
	3								
N = 86	Cree	-0.511	0.765	*	*			*	
	4								
N = 31	Chip	-0.446	0.828	*	*			*	
	5								
N = 37	Other	0.406	0.996			*	*		
RAVEN'S									
	1	-0.025	0.940						0.0033
	2	0.237	0.696						
	3	-0.224	1.060					*	
	4	-0.251	1.067						
	5	0.376	0.851			*			
L-T (V)	1	0.198	1.022			*	*		0.000
	2	0.415	0.776			*	*		
	3	-0.561	0.774	*	*			*	
	4	-0.479	0.857	*	*			*	
	5	0.359	1.049			*	*		
L-T (NV)	1	0.102	0.947			*			0.000
	2	0.317	0.894			*	*		
	3	-0.395	0.932	*	*			*	
	4	-0.404	0.988		*			*	
	5	0.376	0.989			*	*		

* = .05 Level of Significance

do less well on academic and standardized intelligence test, when their performance is compared to that of other children in this population. The native children's abilities on the Raven's appear to be more widely distributed than are the abilities of other students. The standard deviation from the mean for both groups 3 and 4 is significantly greater than that of either groups 1, 2, or 5. The French and other language background students appear to do as well as their unilingual peers on the tests. These results appear to be consistent with conclusions reached by other studies reviewed in Chapter 2.

Correlations

In this section, the correlations between the group intelligence tests and the WISC-R, and C.T.B.S., will be summarized. The correlations with the WISC-R, especially in the specific groups, have to be treated with caution, as the N size of each group is quite small, ranging from 7 to 27.

The data presented in Table 4 indicates the Raven's correlates positively with the other measures. The extent of the correlations ranges from a low of .108, with the WISC-R (V) for group 5, to a high of .826, with the WISC-R (P) for group 3. Also for group 3, it correlates quite highly with the WISC-R (V) (r=7.22) and with the WISC-R (P) (r=.826). In the total group correlations, the Raven's correlates lowest with the WISC-R(V) with an r of .453, and highest with the L-T(NV) with an r of .695. The correlation between the Raven's and the academic achievement measure, the C.T.B.S., is a moderate .550.

The L-T(V) correlates highest with academic achievement in comparison with the other intelligence measures. It has an r of .869 across all groups with the C.T.B.S., the lowest r is .783 for group 2, and the

TABLE 4

Correlations Between Measures

TEST	GROUP	CTBS	RAVEN'S	L-T(V)	L-T(NV)	WISC-R(V)	WISC-R(P)
		(Based on an N of 332)				(Based on an N of 77)	
CTBS	1	1.000	0.544	0.887	0.714	0.793	0.539
	2	1.000	0.392	0.783	0.641	0.578	0.543
	3	1.000	0.593	0.802	0.696	0.884	0.735
	4	1.000	0.468	0.821	0.647	0.688	0.498
	5	1.000	0.671	0.860	0.755	0.944	0.591
	T	1.000	0.550	0.869	0.725	0.745	0.562
RAVEN'S	1	0.544	1.000	0.529	0.684	0.539	0.614
	2	0.392	1.000	0.279	0.548	0.232	0.442
	3	0.593	1.000	0.612	0.750	0.722	0.826
	4	0.468	1.000	0.589	0.725	0.257	0.676
	5	0.671	1.000	0.672	0.686	0.108	0.505
	T	0.550	1.000	0.543	0.695	0.453	0.624
L-T(V)	1	0.887	0.529	1.000	0.762	0.837	0.671
	2	0.783	0.279	1.000	0.625	0.644	0.627
	3	0.802	0.612	1.000	0.787	0.678	0.445
	4	0.821	0.589	1.000	0.776	0.752	0.713
	5	0.860	0.672	1.000	0.865	0.826	0.438
	T	0.869	0.543	1.000	0.780	0.771	0.583
L-T(NV)	1	0.714	0.684	0.762	1.000	0.697	0.788
	2	0.641	0.548	0.625	1.000	0.409	0.614
	3	0.696	0.750	0.787	1.000	0.660	0.671
	4	0.647	0.725	0.776	1.000	0.445	0.653
	5	0.755	0.686	0.865	1.000	0.583	0.668
	T	0.725	0.695	0.780	1.000	0.632	0.704

highest is .887 for group 1. The lowest correlation for the L-T(V) is with the Raven's with an r of .543. The L-T (NV) and WISC-R(V) correlate quite highly with the L-T(V) with correlation coefficients of .780 and .771 respectively.

The L-T (NV) correlations are all consistently high, ranging from .632 with the WISC-R(V), and .780 with the L-T(V).

Considering the non-random method of selecting students for WISC-R assessments, the WISC-R(V) correlates highly with achievement for this population with an r of .745. The WISC-R(P) correlates with the C.T.B.S. more modestly with an r of .562.

A number of general trends become evident in the data in Table 3. The Raven's has consistently lower correlations with other measures than the L-T tests. Another trend in the data, that was not predictable from the literature was the consistantly low correlations between measures for group 2. Of the fourteen correlations between measures for each group, group 2 had 11 instances of the lowest correlations of any group, the second lowest once, and the third twice. The highest correlation for group 2 was an r of .783 between the L-T(V) and C.T.B.S. Their lowest correlation was between the Raven's and the WISC-R(V) with an r of .232. The data also reveals that the L-T(V) is the best predictor of achievement, of all of the intelligence tests for all of the groups. The results appear to support MacArthur's (1966) conclusion that standard intelligence tests are at least as adequate as culture fair tests in predicting academic achievement for Metis and Indians.

Crosstabulations

Looking at the total crosstabulations for Raven's and L-T(V) in Table 5, one can see that of the 332 students, The Raven's and the

TABLE 5

Crosstabulations of Classified Students Amont the Raven's, L-T(V), L-T(NV), and L-T(FS) for Groups and Total Group.

GROUP		L-T(V)			L-T(NV)			L-T(FS)		
	RAVEN'S	QA	NSD	UA	QA	NSD	UA	QA	NSD	UA
1-English	QA	1	24	1	9	17	0	3	23	0
N = 111	NSD	1	69	0	5	58	7	2	67	1
	UA	0	14	1	0	12	3	0	13	2
		r = .079			r = .376			r = .269		
		Sig.= .2045			Sig.= .0000			Sig.= .0021		
2-French	QA	0	11	1	1	11	0	1	11	0
N = 67	NSD	1	43	1	5	37	3	0	43	2
	UA	0	9	1	0	8	2	0	9	1
		r = .006			r = .191			r = .222		
		Sig.= .4794			Sig.= .0605			Sig.= .0358		
3- Cree	QA	0	7	0	1	6	0	1	6	0
N = 86	NSD	4	58	0	3	56	3	2	59	1
	UA	0	17	0	0	13	4	0	16	1
		r = .050			r = .295			r = .198		
		Sig.= .3242			Sig.= .0029			Sig.= .0336		
4- Chip	QA	0	4	0	2	2	0	1	3	0
N = 31	NSD	0	19	0	2	17	0	0	19	0
	UA	0	7	1	0	5	3	0	7	1
		r = .256			r = .573			r = .417		
		Sig.= .0779			Sig.= .0004			Sig.= .0098		
5 - Other	QA	1	1	0	0	2	0	1	1	0
N = 37	NSD	1	29	0	3	25	2	1	28	1
	UA	0	0	5	0	5	0	0	5	0
		r = .325			r = .014			r = .241		
		Sig.= .0248			Sig.= .4673			Sig.= .0752		
TOTAL	QA	2	47	2	13	38	0	7	44	0
N = 332	NSD	7	218	1	18	193	15	5	216	5
	UA	0	52	3	0	43	12	0	50	5
		r = .076			r = .320			r = .249		
		Sig.= .0829			Sig.= .0000			Sig.= .0000		

L-T(V) didn't agree on the classification of 109. The Raven's also disagreed with the L-T(NV) in 114 cases and with the L-T(FS) 104 times. The correlations reflect the low degree of agreement between classifications, as they vary from a high of .573, between the Raven's and the L-T(NV), for group 4, to a low of .006 between the Raven's and L-T(V), for group 2. The correlations for the total groups are also very low, with a low r of .076 between the Raven's and L-T(V) and a high r of .320 between the L-T(NV) and the Raven's. Considering the large number of classification disagreements, the crosstabulations show only two cases of total disagreement. In both cases, when the Raven's was used, the individuals were classified as OAs, whereas when the L-T(V) was used, they were classified as UAs. Scattergrams indicating the extent of agreement between the Raven's and L-T(FS) for each group, and total groups, are located in Appendix B. The crosstabulations in Table 6, where only the WISC-R(FS), L-T(FS), and Raven's are compared, presents a similar pattern to the one observed in Table 5. The correlations between classifications for each group range between .091 for group 5 to .681 for group 3. The exception to this low to moderate agreement between classifications, is the perfect agreement between the L-T(FS) and WISC-R(FS) on the classification of the 7 students in group 5. Table 7 presents the series of crosstabulations between the L-T(V) and (NV) Batteries, the WISC-R(V) and (P) Batteries, and the Raven's classifications. The crosstabulations lend emphasis to the general trend indicating that there is a limited amount of agreement between the instruments in their respective classifications. The tables reveal that the Raven's correlates more closely to the L-T(NV), than it does to the L-T(V) or L-T(FS) for all groups. The data further indicates that both

TABLE 6

Crosstabulations for the Classified Students Among the Raven's, L-T(FS), and WISC-R(FS) for Groups and Total Group.

GROUP		L-T(FS RAVEN'S			WISC-R(FS) RAVEN'S			WISC-R(FS) L-T(FS)		
		QA	NSD	UA	QA	NSD	UA	QA	NSD	UA
1, English N = 27	QA	1	5	0	0	6	0	0	1	0
	NSD	0	13	1	0	9	5	0	17	6
	UA	0	5	2	0	4	3	0	1	2
		r = .414			r = .317			r = .305		
		Sig.= .0159			Sig.= .0539			Sig.= .0627		
2, French N = 16	QA	1	5	0	2	4	0	1	0	0
	NSD	0	7	0	0	6	1	1	13	1
	UA	0	3	0	0	3	0	0	0	0
		r = .289			r = .364			r = .565		
		Sig.= .1389			Sig.= .0828			Sig.=.0113		
3, Cree N = 16	QA	0	1	0	0	1	0	0	2	0
	NSD	2	7	0	0	9	0	0	10	4
	UA	0	6	0	0	2	4	0	0	0
		r = .203			r = .681			r = .218		
		Sig.= .2259			Sig.= .0018			Sig.= .2084		
4, Chip N = 11	QA	1	2	0	0	3	0	0	1	0
	NSD	0	4	0	0	4	0	0	8	1
	UA	0	3	1	0	2	2	0	0	1
		r = .538			r = .541			r = .553		
		Sig. =.0439			Sig.= .0429			Sig.= .0389		
5, Other N = 7	QA	0	1	0	0	1	0	1	0	0
	NSD	1	3	0	1	3	0	0	6	0
	UA	0	2	0	0	2	0	0	0	0
		r = .091			r = .091			r = .1.000		
		Sig.= .4228			Sig.=.4228					
TOTAL N = 77	QA	3	14	0	2	15	0	2	4	0
	NSD	3	34	1	1	31	6	1	54	12
	UA	0	19	3	0	13	9	0	1	3
		r = .313			r = .409			r = .419		
		Sig.= .0028			Sig.= .0001			Sig.= .0001		

TABLE 7

Crosstabulations of Classified Students on the L-T(V) and (NV), WISC-R(V) and (P), and the Raven's for the Total Group.

L-T (V)

Raven's	OA	NSD	UA
OA	1	14	2
NSD	2	36	0
UA	0	21	1

r = .0

Sig.= .5000

L-T(NV)

Raven's	OA	NSD	UA
OA	5	12	0
NSD	7	27	4
UA	0	14	8

r = .427

Sig.= .0001

WISC-R(V)

Raven's	OA	NSD	UA
OA	3	13	1
NSD	3	30	5
UA	0	16	5

r = .219

Sig.= .0279

WISC-R(P)

Raven's	OA	NSD	UA
OA	3	12	2
NSD	0	27	11
UA	0	9	13

r = .428

Sig.= .0001

WISC-R(V)

L-T(V)	OA	NSD	UA
OA	0	3	0
NSD	7	53	11
UA	0	3	0

r = .0

Sig.= .5000

WISC-R(P)

L-T(V)	OA	NSD	UA
OA	0	3	0
NSD	3	44	24
UA	0	1	2

r = .174

Sig. = .0656

WISC-R (V)

L-T (NV)	OA	NSD	UA
OA	3	9	0
NSD	4	43	6
UA	0	7	5

r = .387

Sig. = .0003

WISC-R (P)

L-T(NV)	OA	NSD	UA
OA	3	8	1
NSD	0	38	15
UA	0	2	10

r = .521

Sig. .0000

the Raven's and the L-T(NV) identify a greater proportion of UAs than do the L-T(V) or L-T(FS).

As the WISC-R (FS) is used as the intelligence measure in the screening and placement of students, the students it identified as UAs were compared to those identified by the Raven's and the L-T (FS). Table 6 indicates that of the 15 students the WISC-R(FS) classified as UAs, the L-T(FS) identified 3 and the Raven's identified 9. Therefore, when the group measures were the only instruments used to identify students as UAs, the L-T(FS) would have resulted in 12 Beta errors, and the Raven's would have resulted in 6 Beta errors. The L-T(FS) identified 1 and the Raven's identified 13 students as UAs that the WISC-R(FS) had classified as NSD. The results of the crosstabulations suggest that each intelligence measure has classified the students somewhat differently. The degree of concensus, though generally quite good, does not appear to overlap sufficiently to allow for accurate prediction of inclusion in a given category, from one measure to another. The 3 scattergrams in Appendix C indicate the extent of the agreement between the Raven's, L-T(FS), and WISC-R(FS) graphically.

CHAPTER 5

Discussion, Conclusions and Recommendations

The literature indicated that most of the standard intelligence tests are well constructed, statistically reliable instruments. This study was undertaken to investigate the usefulness of two of these tests rather than their technical adequacy. The two tests, the L-T and Raven's, were administered to a population of students from varied language backgrounds. The test results have been analyzed to determine their utility as contributors in the identification of UAs. This chapter will discuss the results of the study in the context of this objective. The conclusions, and the resulting recommendations will then be presented. Suggestions for further research will conclude this study.

Discussion of the Results

The Lorge-Thorndike Test:

The results of this study tend to substantiate the claim of the authors of the L-T, that it measures mental abilities necessary for academic achievement. The high correlations between the L-T (V) and the C.T.B.S. for all groups indicate the L-T is an adequate predictor of academic success as measured by the C.T.B.S. The results of this study also tend to agree with prior studies that indicated that Natives perform better on the nonverbal tests than they do on the verbal tests. The students who come from French Canadian (group 2) and other language backgrounds (group 5) provided support for the conclusions of similar studies, such as those of Peal and Lambert (1962), Cummings (1974), and Moss (1979), indicating that intelligence test performance is not necessarily hindered by a bilingual environment.

The above paragraph atests to the consistency of the L-T results between studies. An example of this agreement is that both this study, and that of Ellis (1969), found that the L-T and WISC-R do not identify the same population when they are used to identify UAs. The lack of agreement between the L-T and WISC-R in classifying individuals detracts from its usefulness as a gross screening device. When the L-T was used to identify UAs, 80% remained unidentified, due to beta error. This, in turn, decreases the justification for administering the test.

The high degree of correlation between the L-T and C.T.B.S. (r= .819) may also be a basis for questioning its usefulness. The high correlation provokes the question as to whether time and effort should be taken to administer both. Critics of these intelligence tests, such as Williams (1976) and Mercer (1979) claim that low scores on both achievement and intelligence measures have a cumulative detrimental effect on an individual's education. They feel that teachers invariably lower their expectations for students who have low scores, without actually changing the program of studies. They are of the opinion that this results in the establishment of a self-fulfilling syndrome. The circular reasoning that arises is that the low achievement score of an individual is explained by the low intelligence result, and the low intelligence result is the justification for setting low achievement goals. In order to avoid this self-fulfilling prophesy, Boozer (1976), Flaugher (1978) and Deutsch (1979) among others, are strong advocates of criterion referenced testing and the de-emphasizing of norm referenced tests of all types. They are of the opinion that well constructed measures for monitoring the development and progress of a child should

eliminate the needless and sometimes harmful reference to other students or groups.

The Standard Progressive Matrices:

The Raven's results were also similar to the results found in other studies. They indicated that the Raven's measured mental abilities, or a general intelligence factor, that was generally more evenly distributed across racial and ethnic groups. The results of this study indicated that the Metis and Indians in most cases did not differ significantly in their scores on this test as compared to the other groups. The results also indicated that the Raven's correlated moderately with all the other measures. In spite of only moderate correlations with the WISC-R, the Raven's identified 3 times as many UAs as did the L-T. Part of the reason for the moderate correlation with the WISC-R was that the Raven's identified 13 students as UAs that the WISC-R had classified as NSD.

There appears to be a difference of explanation for the low correlations between the L-T and WISC-R, in their crosstabulation classifications, and the Raven's and the WISC-R (FS) in their classifications. The L-T appeared to agree with the WISC-R (FS) 59 out of 77 times, yet identified only 3 UAs in agreement with the WISC-R. The Raven's agreed with the WISC-R 42 times, and identified 9 UAs in agreement with the WISC-R. The impression left by these results is that the L-T is a test that is similar to the WISC-R in what it intends to measure, but that it is less accurate. The Raven's on the other hand appears to be an instrument that differs in focus from the WISC-R, yet it measures an area of intelligence common to both.

The L-T has a higher correlation with the C.T.B.S. than does either the WISC-R, or the Raven's. It appears that both the WISC-R and the Raven's measure intellectual abilities that are not presently as directly related to academic achievement as are the abilities measured by the L-T.

The Minority Groups:

Though the Raven's appeared to be more culture-fair in assessing the intelligence of Native children, both the Metis and Indians still had lower means on it, as they did on all of the measures. As a result, the Natives in this study would be the ones that could be most detrimentally affected by the previously mentioned circular reasoning syndrome. Bowd (1977), in his study of the educational policies with regard to Indians, concluded that there is a general dissatisfaction in native communities across Canada with the present educational programs. Titly (1981), points out that the Natives, feeling that the regular school curriculums, standards and assessments procedures are inappropriate, are attempting to establish their own schools. If the results of this study concur with the results obtained in the past, their concerns may be justified.

The performance of the students in group 2, with a French language background, has produced both expected, as well as not readily explainable, results. The literature that was reviewed indicated that these students generally perform at the average to above average levels on intelligence and achievement tests. This study tends to support this conclusion. The aspect of the results that is less readily understandable is that their correlations between measures are generally lower than those of the other groups. There are numerous possible contributing

factors to this, one plausable factor is that the large percentage of students in this group are conversant in their parents' language. Though not a part of this study, an informal survey revealed that 46% (31 out of 67) of the students in this group spoke French at home. This meant that there were virtually two distinct groups in the French Canadian background catagory. This may have resulted in less consistent results across measures. Another possible factor is that the groups were formed across grades. Kittel's (1963) study led him to recommend caution when assessing younger French bilinguals. He felt that there occurred vacillation between one and the other language in their early years, which would diminish the consistency of their results. If this movement back and forth between languages occurred in this study, it would decrease the amount of agreement between the measures.

There are other equally plausable postulates that can account for the lower correlations for group 2. These answers however, are not readily evident, and merit further study.

The issue of bias has been discussed in this study. As Lorge (1966) points out, when tests indicate differences between individuals and groups, that does not necessarily mean that the tests are biased against the lower scoring individuals, or groups. Vernon (1979) points out in his discussion on overrepresentations in special classes, that in the case of overrepresentation of specific minorities in certain sports, none of the minorities consider this a case of bias. It appears in this study that the tests that have been evaluated are reliable, and unbiased instruments. It is in the use of these results, that the problem of bias may enter. That is, if the results are used to justify inferior education, or are overinterpreted to generalizations about

classes of people,or are ignored, then, at that stage, bias can be considered to have entered into the assessment procedure.

Conclusions

This study has apparently raised as many questions as it has answered. Some of the relevant questions are as follows: Would the between group results in specific grades differ at different grade levels ? Is there a significant difference in the results of the three schools ? Is there a difference in the results of the Metis who live on the settlement and the Indians who live on the reserve, as compared to the Metis and Indians who live in the towns ? What would the correlation and crosstabulation results be if everyone in the population had been assessed on the WISC-R, or if, at least a random sample had been assessed ? Most of these questions require larger populations or careful longitudinal study before they can be satisfactorily answered.

The study has also raised questions to issues that can, and perhaps should be answered in the near future. How well do teachers and administrators understand the uses and limitations of assessment measures and their results ? How much emphasis do the results receive in comparison to other information available on the student ? Are there appropriate programs and materials, trained teachers, and support personnel, to ensure that assessment results culminate in positive meaningful actions ? Are the educational programs based on an individual's strengths, or are they focussed on his weaknesses ? These questions require fairly immediate answers. The initial purpose of this study was to examine the adequacy of the L-T as an instrument used in the identification of UAs. The answer to that appears quite evident, as the results indicated that 80% of those classified by the WISC-R

as UAs were not identified on the L-T due to beta error.

The second purpose was to question the adequacy of the Raven's in identifying UAs. The answer here is less evident, as it appears that at the present time, academic achievement is not very closely related to that area of intelligence measured by the Raven's. However, the continued use of the Raven's could allow students classified as low achievers - low intelligence to be viewed in a different light. With a change in perspective, educational programs may begin to develop in a direction that taps these mental abilities. This could result in students experiencing success through their strengths, finding more satisfaction and meaning in their school work, and taking full advantage of their educational experiences. The usefulness of the Raven's would therefore have to be considered in the light of possible use, rather than present use.

The conclusions reached on the assessment instruments are restricted to the scope of this study. This study, primarily deals with underachievement, therefore, the adequacy or usefulness of the instruments in other situations, or for other purposes, has not been examined. This study however, seriously questions the relative benefits of group assessments, as compared to the possible negative consequences that might result, especially for Native children.

Recommendations

The results of this study indicate that the issue of whether or not to group test is a serious question. It appears that there is a definite need for dialogue on the advantages of possible negative consequences of group testing in general, and of group intelligence testing in particular. If, after weighing the benefits and consequences

of assessment, the decision is reached to continue with group assessments, perhaps more recent tests should be evaluated, before continuing with the use of the L-T. An obvious alternative is a recently modified version of the L-T called the Canadian Cognitive Abilities Test (C.C.A.T.). It was developed by Thorndike and Hagen who are two of the authors of the L-T. It is normed on the same population as was the C.T.B.S. and adapted for Canadian use by Wright, the same person who adapted the L-T for Canadian use. (Thorndike and Hagen, 1971).

An issue related to the group testing question is the continuing practice of the recording of a group intelligence test result in an individual's commulative record. The intent of this practice is to give the user of the record an indication of the individual's intellectual ability. It has yet to be established that a group test score is an effective way of relating meaningful and relevant information about the intellectual abilities of a given individual. The recommendation is, that the possible negative consequences, and limited benefits of this practice, warrant its re-evaluation.

The individual assessments for students grades 3 through 8 with a WISC-R, appear to be the best method available at the present time for assessing an individual's intelligence. That is not to imply that these tests are flawless. As these tests are not flawless, it is recommended that an ongoing monitoring system, of the intelligence measurement field, should be instituted. Developments in this field can then be readily evaluated in terms of their relevance for a given region. A modification that this study would recommend is that the Raven's should accompany the WISC-R as part of the assessment procedure. The effects of this could then be evaluated through the monitoring system, especially

with regard to Indian and Metis children. Although this recommendation would necessitate the development of local norms, it would allow the student to be viewed from another perspective with regard to their intellectual abilities.

Serious consideration should be given to the establishment of an evaluation system based on criterion referenced measures. That may entail workshops and the inservicing of personnel over a period of time. The benefits would include; the creation of a pool of items measuring numerous skills, the development of hierarchical progressions of skill and knowledge levels in various subject areas, the use of content and materials that are more appropriate for the students of the area, and possible experimentation with innovative educational approaches that may prove to be more effective than the traditional methods. Additional benefits would result in having personnel who were knowledgeable as to the content, goals, and structure of the curriculum, who were trained in the development of reliable and appropriate test items, and who were intimately involved and had a personal investment in the development of the assessment procedures.

In summary, it would appear that the days of educators making unquestioned decisions with regard to tests, placement of students, programs of studies, or promotions of students, are quickly coming to an end. The more self-critical the educational community is of its testing procedures now, the less negative will be the public's reaction in the future.

Suggestions for Further Research

The areas that warrant further research were presented earlier in the study in the form of questions yet to be answered. They may be

summarized as follows:

1. More recent group tests need assessing as to their adequacy in the identification of underachievers. Ideally, these studies should be carried out with a population larger than the one in this study, enabling both between group and between grade analysis of the data. A larger population would also allow for an investigation into the question of whether true bilinguals differed from individuals whose parents spoke another language, in their response sets and test results.
2. As intelligence test results are recorded in cummulative files, an area that needs investigation is the extent to which educators have knowledge of what intelligence tests measure, their strengths, and their limitations. An additional area of research would be to ascertain the amount of importance placed on these results by educators.
3. Research is needed into the results obtained when the WISC-R is administered randomly, or to a whole population. These results could then be correlated with group test results and also compared to teachers' observations, to determine the most effective way of identifying underachievers.
4. Research of the use of the Raven's for this population is necessary in order to determine whether the published norms are appropriate, and the educational outcome of combining the Raven's with the WISC-R in the assessment procedure.

REFERENCES

References

Adelberto, A., Intelligence Testing and the Quality of Life. Educational Research Quarterly,]979, 4, 3 -]2.

Akhurst, B.A., Assessing Intellectual Ability. New York: Burnes and Noble Publishing Co.,]970.

Anastasi, A., On the Formation of Psychological Traits. American Psychologist,]970, 25, 899 - 9]0.

Anastasi, A., Psychological Testing. (3rd. ed.) New York: MacMillan Publishing Co.,]976.

Barnes, E., I.Q. Testing and Minority Children: Imperatives for Change. Journal of Non - White Concerns,]973, 2, 4 -]9.

Bersoff, D., Silk Purses Into Sows Ears: The Decline of Psychological Testing and a Suggestion for Its Redemption. American Psychologist, 1973,]0, 892 - 899.

BITCH Test (Black Intelligence Test of Cultural Homogeneity. in O.K. Buros (Ed.) Intelligence Tests and Reviews: Seventh Mental Measurement Yearbook, Highland Park, N.J.: Gryphon Press,]975, 343.

Bloom, B.S., Stability and Change in Human Characteristics. New York: Wiley and Sons,]964.

Boozer, B., An Alternative to Intelligence Testing for Minority Children. The Journal of Negro Education,]978, 47, 4]4 -4]8.

Bortner, M., Review of the Raven's. in O.K. Buros (Ed.) Intelligence Tests and Reviews: Seventh Mental Measurements Yearbook, Highland Park, N.J.: Gryphon Press,]975, 695 - 696.

Boyd , A.D., Ten years After the Hawthorn Report: Charging for the Education of Canadian Native Peoples. Canadian Psychological Review, 1 977,]8, 332 - 345.

Bryan, T.H. and Bryan, J.H., Understanding Learning Disabilities. (3rd ed.) Sherman Oaks, California: Alfred Publications, 1978.

Cardinal, H., The Unjust Society, Edmonton, Alberta: Hurtig Publishing Co., 1969.

Carlson, J.S. and Wiedl,K.H. Dynamic Assessment: An Approach Toward Reducing Test Bias. Paper presented at the annual meeting of the Western Psychological Association, Honolulu: May, 1980.

Cattell, R.B. Handbook for the Culture Fair Intelligence Test: A Measure of "g". Champagne, Ill.: Institute for Personality and Ability Testing, 1959.

Cattell, R.B., Theory of Fluid and Crystallized Intelligence: A Critical Experiment. Journal of Educational Psychology, 1963, 54,1 - 22.

Chismar, Sister M.H., A Study of the Effectiveness of Cross-Level Grouping of Middle School Underachievers for Reading Instruction. Unpublished Ph.D. Thesis, Kent State, 1971.

Clarizio, H.F., In Defence of the I.Q. Test. School Psychology Digest, 1979, 8, 79 - 88.

Cleary, T.A., Humpherys, L.G., Kendrick, S.A., and Wesman, A., Educational Uses of Tests With Disadvantaged Students. American Psychologist, 1975, 30, 15 -41.

Cohen, J.S., Legal Issues in Non Discriminatory Testing. School Psychology Digest, 1977, 6, 35 - 42.

Conwell, W.L., The Inheritance of I.Q. and Scholastic Achievement: Further Comments on the Jensen Article. The Journal of Negro Education, 1980, 49, 97 - 104.

Cottance, T., Cumulative Teacher Effect: Do Schools Consistantly Influence the Performance of Their Students? Educational Review, 1980, 30, 267 - 280.

Crawford, C., George Washington, Abrahm Lincoln and A.Jensen: Are They Compatible? American Psychologist, 1979, 34, 664 - 672.

Cress, J.W., Cognitive and Personality Testing Use and Abuse. Journal of American Indian Education, 1974, 13, 16 - 19.

Cronbach, L.J., Five Decades of Public Controversy Over Mental Testing. American Psychologist, 1975, 30, 1 - 14.

Cummins, J.P., Bilingualism and Cognitive Representation. Unpublished Ph. D. Thesis University of Alberta, 1974.

Cundick, B.P., Measures of Intelligence on Southwest Indian Students. The Journal of Social Psychology, 1970, 81, 151 - 156.

Dean, R.S., Predictive Validity of the WISC-R With Mexican - American Children. Journal of School Psychology, 1979, 17, 55 - 66.

Dean, R.S., Factor Structure of the WISC-R With Anglos and Mexican - Americans. Journal of School Psychology, 1980, 18, 234 - 239.

deAvila, E.A. and Havassy, B., I.Q. Testing and Minority Children. Austin, Texas: Dissemation Center for Bilingual and Bicultural Education, January, 1974.

Deutsch, M., Education and Distributive Justice: Some Reflections on Grading Systems. American Psychologist, 1979, 34, 5391 - 5401.

Ebel, R.L., The Social Consequences of Educational Testing. in A. Anastasi (Ed.) Testing Problems in Perspective, Washington, D.C.: American Council on Education, 1966, 18 - 28.

Elley, W.B. and MacArthur, R.S., The Standard Progressive Matrices as a Culture Reduced Measure of General Intellectual Ability. Alberta Journal of Educational Research, 1962, 8, 54 - 65.

Ellis, J.R. Variables Related to the Identification of Underachievers; Final Report. Washington, D.C.: Department of Health, Education and Welfare, September, 1969.

Ewanyshyn, E., The Evaluation of a Ukrainian Bilingual Program. Unpublished M. Ed. Thesis, University of Alberta, 1978.

Fine, B., The Stranglehold of the I.Q. Garden City, New York: Doubleday Publishing Co., 1975.

Fisk, B., Identifying Learning Disabilities. Unpublished M.Ed. Thesis, University of Alberta, 1979.

Flaugher, R.L., The Many Definitions of Test Bias. American Psychologist, 1978, 33, 671 - 679.

Fuller, R., In Search of the I.Q. Correlation. Stoney Brook, New York: Ball - Stick - Bird Publishing Co., 1977.

Goodman, J.J., Is Tissue the Issue? A Critique of SOMPA's Models and Tests. School Psychology Digest, 1979, 8, 47 - 62.

Goslin, D.A., The Search for Ability: Standardized Testing in Perspective. New York: Wiley and Sons, 1963.

Green, B.T., In Defence of Measurement. American Psychologist, 1978, 33, 64 - 70.

Greenleaf, W. and Smith, H.L., Non-Discriminatory Intellectual Assessment Procedures. Baton Rouge, Louisiana State Department of Education, 1978.

Gronbind, N., Measurement and Evaluation in Teaching. (3rd ed.) New York: MacMillan Publishing Co., 1976.

Grueberg, H. and Opper, S., Piaget's Theory of Intellectual Development. (2nd ed.) Englewood Cliffs, N.J.: Prentice Hall Ltd., 1979.

Guilford, J.P., The Nature of Human Intelligence. New York: McGraw-Hill Publishing Co., 1967.

Guilford, J.P., The Analysis of Intelligence, New York: McGraw-Hill Publishing Co., 1971.

Guilford, J.P., Fluid and Crystallized Intelligence: Two Fanciful Concepts. Psychological Bulletin, 1980, 88, 406 - 412.

Gutkin, T.B. and Reynold, C.R., Factor Similarities of the WISC-R for Anglos and Chicanos Reffered for Psychological Services. Journal of School Psychology, 1980, 18, 34 - 39.

Havinghurst, R.J., Mental Development and School Achievement of American Indian Children and Youth: Series IV, #3, Final Report. Washington, D.C.: Office of Education, June , 1970, BR - 8-0147.

Hays, J.R. and Smith, A.L., Comparison of WISC-R and Culture Fair Intelligence Test Scores for Three Ethnic Groups of Juvenile Delinquents. Psychological Reports, 1980, 46, 931 - 934.

Heath, R.W. and Nielson, M.A., Testing: Evaluation and the Indian Education Act of 1972. Washington,D.C.: Office of Indian Education, April, 1973.

Hebb, D.O., A Textbook of Psychology. (2nd ed.) Philadelphia: Sanders Publishing Co., 1966.

Hieronymus, A.N. and King, E., (Eds.) Canadian Tests of Basic Skills: Manual for Administrators, Supervisors and Counsellors. Thomas Nelson and Sons (Canada) Ltd., 1975.

Hobbs, N., The Future of Children: Categories, Labels and Their Consequences. San Francisco: Jossey - Bass Publishers, 1975.

Horn, J.L. and Cattell, R.B., Refinement and Testing of the Theory of Fluid and Crystallized Intelligence. Journal of Educational Psychology, 1966, 57, 253 - 270.

Hunt, J.McV., Intelligence and Experience. New York: Ronald Press, 1961.

Hynd, G.W. and Garcia, W.I., Intellectual Assessment of the North American Student. School Psychology Digest, 1979, 8, 446 - 454.

Jensen, A.R., How Much Can We Boost I.Q. and Scholastic Achievement? Harvard Educational Review, 1969, 39, 1 - 123.

Jensen, A.R., Another Look at Culture-Fair Testing. in J. Hellmoth (Ed.) Disadvantaged Child: Vol. 3, New York: Bruner/Mazel Publishers, 1970, 53 - 101.

Jensen, A.R., Level I and Level II Abilities in Three Ethnic Groups. American Educational Research Journal, 1973, 10, 263 - 276.

Jensen, A.R., Let's Understand Skodak and Skeels, Finally. Educational Psychologist, 1973a, 10, 30 - 35.

Jensen, A.R., Cumulative Deficit: A Testable Hypothesis? Developmental Psychology, 1974, 10, 996 - 1019.

Jensen, A.R., Test Bias and Construct Validity. Paper presented at the Annual Meeting of the American Psychological Association, Montreal: September, 1975.

Jensen, A.R., The Meaning of Heritability in the Behavioral Sciences. Educational Psychologist, 1975a, 11, 171 183.

Jensen, A.R., Cummulative Deficit in I.Q. of Blacks in the Rural South. Developmental Psychology, 1977, 13, 184 - 191.

Kamin, L.J., Social and Legal Consequences of I.Q. Tests as Classification Instruments: Some Warnings From the Past. Journal of School Psychology, 1975, 13, 317 - 323.

Kaplan, A., The Conduct of Inquiry: Methodology for Behavioral Science. San Francisco: Chandler Press, 1964.

Keech, C.C., Discrepencies Between the Stanford-Binet and Three Group Tests of Intelligence in the Identification of High I.Q. Children. Dissertation Abstracts international, 27:1660A, 1966.

Kehn,C., Review of the WISC-R. Journal of Educational Measurement, 1975, 12, 140 - 144.

Kirkland, M.C., The Effects of Tests on Students and Schools. Review of Educational Research, 1971, 41, 303 - 350.

Kittell, J.E., Intelligence Test Performance of Children from Bilingual Environments. Elementary School Journal, 1963, 64, 76 - 83.

Kundel, A., Discrepencies Between the Stanford-Binet and Three Group Tests of Intelligence in the Identification of Low I.Q. Children. Dissertation Abstracts International, 27:1659A, 1966.

Lambert, W.E. and Tucker, G.R., The Benefits of Bilingualism. Psychology Today, 1973, September, 89 - 94.

Lampert, N.R. Legal Challenges to Testing - Larry P.: A Case in Point. Paper presented at the Annual Meeting of the Aperican Psychological Association, August, 1978.

Liedke, W.W. and Neilson, L.D., Concept Formation and Bilingualism. Alberta Journal of Educational Research, 1968, 14, 225 - 232.

Long, P.A. and Anthony, J.J. The Measurement of Retardation by a Culture - Specific Test. Psychology in the Schools, 1974, 11, 310 - 312.

Lorge, I., Difference or Bias in Tests of Intelligence. in A. Anastasi (Ed.) Testing Problems in Perspective, Washington, D.C.: American Council on Education, 1966, 465 - 471.

Lorge, I., Thorndike, R.L. and Hagen, E., Canadian Lorge-Thorndike Tests: Manual for Administration. Toronto: Thomas Nelson and Sons, 1967.

MacArthur, R.S., Assessing the Intellectual Ability of Indians and Metis Pupils at Ft. Simpson, N.W.T. Ottawa: Canadian Department of Northern and National Resourses, 1962.

MacArthur, R.S. Mental Abilities in a Cross Cultural Context. Paper presented to the Department of Psychology Colloquium, McGill University, Montreal: March, 1966.

MacArthur, R.S., Longitudinal Prediction of School Achievement for Metis and Eskimo Pupils. Paper presented to the C.I.C.R.E. Confrence on Educational Research, Winnipeg: June, 1967.

Macdonald, H.A. and Heatherton, A.H., Contribution of a Non-Verbal General Ability Test to the Educational Assessment of Pupils in the Cross-Cultural Setting of the Canadian North. Journal of Educational Research,1969, 62, 315 - 319.

Mackler, B. and Holman, D. Assessing, Packaging and Delivering: Tests, Tesing and Race. Young Children, 1976, 31, 351 - 364.

Matuzzek, P. and Oakland, T., Factors Influencing Teachers and Psychologists Recommendations Regarding Special Class Placement. Journal of School Psychology, 1979, 17, 116 - 130.

McNemar, Q., Lost: Our Intelligence? Why?. American Psychologist, 1964, 19, 871 - 882.

Mercer, J.R., I.Q.= The Lethal Label. Psychology Today, 1972, September, 44 - 50, 97, 99.

Mercer, J.R., Labeling the Mentally Handicapped. Berkely: University of California Press, 1973.

Mercer, J.R., A Policy Statement on Assessment Procedures and the Rights of Children. Harvard Educational Review, 1974, 44, .25 - 141.

Mercer, J.R., Implications of Current Assessment Procedures for Mexican American Children. Bilingual Education Series, Vol. 1: August, 1977.

Mercer, J.R., In Defence of Racial and Cultural Non-Discriminatory Assessment. School Psychology Digest, 1979, 8, 89 - 115.

Messick, S., Test Validity and the Ethics of Assessment. American Psychologist, 1980, 35, 1012 - 1027.

Miele, F. Cultural Bias in the WISC-R Intelligence Test. American Psychologist, 1979, 34, 149 - 164.

More, A.J. and Oldridge, B.,An Approach to Non-Discriminatory Assessment of Native Indian Children. B.C. Journal of Special Education,1980,4,51-59.

Moss, M., The Effects of Bilinguialism on Cognition. Unpublished M. Ed. Thesis, University of Alberta, 1979.

N.A.A.C.P., Report on Minority Testing. New York: May, 1976.

Nichols, N.J. and McKinney, A.W., Black and White Socio-Economically Disadvantaged People. Journal of Negro Education,1977, 46, 443-449.

Learning Potential.Paper presented at the Annual Meeting of the American Psychological Association, Montreal: September, 1980.

Page, E.B., Decisions for the Handicapped: A New Guide to Evaluation Under the New Laws. Journal of Special Education,1980, 14, 423 - 483.

Peal, E. and Lambert, W.E., The Relationship of Bilingualism to Intelligence. Psychological Monographs, 1962, 76, 546.

Piaget, J., What is Psychology? translated by C. Kumivi, in American Psychologist, 1978, 33, 648 - 652.

Pray, B.S., A Step Toward Fairness in Evaluating Handicapped Indian Children: A Formula for Weighing Cultural Items on the WISC-R and WAIS. Washington, D. C.: Bureau of Indian Affairs Research Bulletin # 1,.Vol. 7, 1979.

Purl, M.C. and Curtic, J., Raven's Progressive Matrices Test Correlations with Measures of Lorge-Thorndike and Wechsler Intelligence Tests. Sacramento; California State Department of Education, 1970, September.

Ramirez, M.M., Taylor, C.Jr. and Petersen, B., Mexican-American Cultural Membership and Adjustment to School. Developmental Psychology, 1971, 4, 141 - 148.

Raven, J.C., Guide to Using the Standard Progressive Matrices. London: Lewis Publishing Company, 1960.

Reschly, D.J., WISC-R Factor Structures Among Anglos, Blacks, Chicanos and Native-American Papagos. Journal of Consulting and Clinical Psychology, 1978, 46, 417 - 422.

Reschly, D.J. and Reschly, J.E., Validity of WISC-R Factor Scores in Predicting Achievement and Attention for Four Sociocultural Groups. Journal of School Psychology, 1979, 17, 355 - 360.

Richmond, B.O. and Long, M., WISC-R and PPVT Scores for Black and White Mentally Retarded Children. Journal of School Psychology, 1977, 15, 261 - 263.

Rosenthal, R., The Pygmalion Effect Lives. Psychology Today, 1973, September, 56 - 63.

Rosenthal, R. and Jacobson, L., Pygmalion in the Classroom. New York: Holt, Rinehart and Winston Publishers, 1968.

St. John, J., Kirchev, A. and Bowman, E., Northwestern Ontario Indian Children and the WISC-R. Psychology in the Schools, 1976, 13, 407-411.

Salvia, J. and Ysseldyke, J.E., Assessment in Special and Remedial Education. (2nd ed.) Boston: Houghton Mifflin Publishing Co., 1981.

Sandaval, J., The WISC-R and Internal Evidence of Test Bias With Minority Groups. Journal of Counseling and Clinical Psychology, 1979, 47, 919 - 926.

Sarasan, B. and Doris, J., Psychological Problems in Mental Deficiency. (4th ed.) New York: Harper and Row Publishing Co., 1969.

Sattler, J.M. Assessment of Children's Intelligence. (Revised) Philadelphia: W. L. Saunders Publishing Co., 1974.

Seyfort, B., Spreen, O. and Lahmer, V., A Critical Look at the WISC-R with Native Indian Children. The Alberta Journal of Educational Research, 1980, 26, 14 - 24.

Skagar, R. and Fitz-Gibbon, C., Mentally Gifted Disadvantaged Students: An Investigation of Methods of Identification, Including the Use of "Culture Fair" Tests at the Grade Eight Level. Final Report. Los Angeles: University of California, April, 1972.

Smith, C.R. and Knoff, H.M. School Psychologists and Special Education Students' Placement Decisions: I.Q. Still Tips the Scale. The Journal of Special Education, 1981, 15, 55 - 64.

Spearman, C., The Abilities of Man. London: MacMillan Publising co., 1932.

Terman, L.M., A Symposium, Intelligence and Its Measurement. Journal of Educational Psycology, 1921, 12, 127 - 133.

Thorndike, R.L., Exhibit C: I.Q. Frail, Defence Experts Testify. A.P.A. Monitor, 1978, 8.

Thorndike, R.L. and Hagen, E., The Canadian Cognitive Abilities Test: Manual for Administrators. Thomas Nelson and Sons (Canada) Ltd., 1971.

Thorndike, R.L. and Hagen, E., Measurement and Evaluation in Psychology and Education. (4th ed.) New York: John Wiley and Sons Ltd., 1977.

Thurstone, L.L. and Thurstone, T., Primary Mental Abilities Test. Chicago: Science Research Associates, 1965.

Titly. B., The Struggle for Local Control. Journal of American Indian Education, 1981, 20, 19 - 24.

Tryon, W.W., The Test-Trait Fallacy. American Psychologist, 1979, 34, 402 - 406.

Tucker, J.A., Ethnic Proportion in Classes for Learning Disabled: Issues in Non Biased Assessment. Journal of Special Education, 1980,14,93-105.

Tuddenham, R.D., The Nature and Measurement of Intelligence. in L.J. Postman (Ed.) Psychology in the Making, New York: Knoff Publishing Co., 1962, 469 - 525.

Vernon, P.E., The Structure of Human Abilities. (2nd ed.) London: Mathuen Publishing Co., 1961.

Vernon, P.E., Ability Factors and Environmental Influences. American Psychologist, 1965, 20, 723 - 733.

Vernon, P.E., Intelligence and Cultural Environment. London: Methuen Publishing Co., 1969.

Vernon, P.E., Intelligence: Heredity and Environment. San Francisco: W.H. Freeman Publishing Co., 1979.

Vygotsky, L.S., Thought and Language. Cambridge: M.I.T. Press, 1962.

Wechsler, D., The Measurement and Appraisal of Adult Intelligence. (4th ed.) Baltimore: Williams and Williams Publishing Co., 1958.

Wechsler, D., Manual for the Wechsler Intelligence Scale for Children - Revised. New York: Psychological Corporation, 1974.

West, L.W., Assessing Intellectual Ability With a Minimum of Cultural Bias for Two Samples of Metis and Indian Children. Unpublished M. Ed. Thesis, University of Alberta, 1962.

West, L.W. and MacArthur, R.S., An Evaluation of Selected Intelligence Tests for Two Samples of Metis and Indian Children. Alberta Journal of Educational Research, 1964, 10, 17 - 27.

Williams, R.L. Black Pride, Academic Relevance, and Individual Achievement. The Counseling Psychologist, 1970, 2, 18 - 22.

Williams, R. L., Abuses and Misuses in Testing Black Children. The Counseling Psychologist, 1971, 1, 62 - 73.

Williams, R.L. and Kirkland, J., The White Counselor and the Black Client. The Counseling Psychologist, 1971, 2, 114 - 116.

Williams, T., Teachers Prophesies and the Inheritance of Inequality. Sociology of Education, 1976, 49, 223 - 235.

Winger, L.J., The Application of A Special Counseling Technique to Maladjusted Under-Achievers: A Pilot Project. Washington, D.C.: Office of Education Final Report August, 1968.

Zirkel, P.A., Spanish-Speaking Students and Standardized Tests. Urban Review. 1972, 5, 32 - 40.

APPENDICES

APPENDIX A

Distribution of Scores for the Various Group Measures by Grade and by Group.

Distribution of C.T.B.S. Raw Scores by Grades and Groups.

GRADE	N	GROUP	-30	30-49	50-69	70-89	90-109	110-129	130-149	150-169	170+
3	27	1	-	3	10	6	2	5	1	-	-
	10	2	-	-	2	3	4	1	-	-	-
	14	3	4	3	3	2	2	-	-	-	-
	7	4	-	3	2	-	2	-	-	-	-
	3	5	-	-	-	2	-	1	-	-	-
TOTAL:	61		4	9	17	13	10	7	1	-	-
4	18	1	-	1	6	-	6	3	1	1	-
	7	2	-	2	1	1	1	1	1	-	-
	18	3	-	5	6	5	2	-	-	-	-
	6	4	-	-	4	-	1	1	-	-	-
	6	5	-	-	2	1	1	-	1	1	-
TOTAL:	55		-	8	19	7	11	5	3	2	-
5	15	1	-	3	3	3	1	3	1	1	-
	8	2	-	-	-	-	4	4	-	-	-
	15	3	-	3	3	3	1	5	-	-	-
	8	4	-	-	5	1	2	-	-	-	-
	9	5	-	-	-	3	2	1	2	1	-
TOTAL:	55		-	6	11	10	10	13	3	2	-
6	15	1	-	1	1	5	4	1	-	1	2
	15	2	-	-	-	4	5	2	3	1	-
	15	3	-	2	4	4	4	4	-	-	-
	5	4	-	-	1	1	1	1	1	-	-
	2	5	-	-	-	-	2	-	-	-	-
TOTAL:	55		-	3	6	14	16	8	4	2	2
7	20	1	-	1	1	4	4	5	1	3	1
	12	2	-	-	1	2	2	5	1	1	-
	15	3	-	-	5	8	-	2	-	-	-
	1	4	-	1	-	-	-	-	-	-	-
	6	5	-	-	-	2	1	2	-	1	-
TOTAL:	54		-	2	7	16	7	14	2	5	1
8	16	1	-	-	3	5	4	2	2	-	-
	15	2	-	-	2	4	3	1	4	-	1
	6	3	-	-	1	4	1	-	-	-	-
	4	4	-	2	-	-	1	1	-	-	-
	11	5	-	1	1	2	2	2	2	1	-
TOTAL:	52		-	3	7	15	11	6	8	1	1

Distribution of L-T (V) Deviation I.Q.'s by Grades and Groups.

GRADE	N	GROUP	-70	70-79	80-89	90-99	100-109	110-119	120-129	130+
3	27	1	-	3	8	6	3	4	1	2
	10	2	-	1	1	2	4	1	1	-
	14	3	5	2	4	1	-	2	-	-
	7	4	2	-	-	2	3	-	-	-
	3	5	-	-	1	-	1	-	1	-
TOTAL:	61		7	6	14	11	11	7	3	2
4	18	1	1	1	2	3	3	3	5	-
	7	2	-	1	-	2	2	1	-	1
	18	3	3	2	6	2	5	-	-	-
	6	4	-	3	-	-	2	1	-	-
	6	5	-	-	-	2	1	1	2	-
TOTAL:	55		4	7	8	9	13	6	7	1
5	15	1	-	3	3	3	3	1	2	-
	8	2	-	-	-	3	3	2	-	-
	15	3	-	5	3	3	3	1	-	-
	8	4	-	4	2	2	-	-	-	-
	9	5	-	-	2	4	-	2	-	1
TOTAL:	55		-	12	10	15	9	6	2	1
6	15	1	-	2	6	3	1	-	1	2
	15	2	-	1	2	4	2	5	1	-
	18	3	2	4	8	3	1	-	-	-
	5	4	1	1	1	-	2	-	-	-
	2	5	-	-	-	1	1	-	-	-
TOTAL:	55		3	8	17	11	7	5	2	2
7	20	1	-	2	3	5	3	5	1	1
	12	2	-	1	1	5	3	2	-	-
	15	3	1	3	5	3	3	-	-	-
	1	4	1	-	-	-	-	-	-	-
	6	5	-	-	1	3	1	1	-	-
TOTAL:	54		2	6	10	16	10	8	1	1
8	16	1	-	3	2	6	5	-	-	-
	15	2	-	1	2	5	4	2	1	-
	6	3	1	-	3	2	-	-	-	-
	4	4	-	2	-	1	1	-	-	-
	11	5	1	2	-	5	1	1	1	-
TOTAL:	52		2	8	7	19	11	3	2	-

Distribution of L-T (NV) Deviation I.Q.'s by Grades and Groups.

GRADE	N	GROUP	- 70	70-79	80-89	90-99	100-109	110-119	120-129	130+
3	27	1	1	6	2	3	5	6	2	2
	10	2	-	2	-	-	3	3	1	1
	14	3	4	3	3	1	2	-	1	-
	7	4	1	-	2	1	1	2	-	-
	3	5	-	-	-	1	1	1	-	-
TOTAL:	61		6	11	7	6	12	12	4	3
4	18	1	-	2	1	3	1	3	6	2
	7	2	-	-	1	1	-	3	2	-
	18	3	2	4	2	3	5	1	-	1
	6	4	1	-	2	1	-	1	-	1
	6	5	-	-	-	-	1	2	-	3
TOTAL:	55		3	6	6	8	7	10	8	7
5	15	1	-	-	3	5	2	3	1	1
	8	2	-	-	-	1	3	3	-	1
	15	3	-	2	3	4	1	5	-	-
	8	4	-	2	4	-	2	-	-	-
	9	5	-	-	-	1	5	1	1	1
TOTAL:	55		-	4	10	11	13	12	2	3
6	15	1	-	2	-	2	2	3	4	2
	15	2	-	1	3	-	4	3	3	1
	18	3	-	4	1	5	4	3	1	-
	5	4	1	1	-	-	2	1	-	-
	2	5	-	-	-	-	1	-	-	1
TOTAL:	55		1	8	4	7	13	10	8	4
7	20	1	-	1	-	1	5	7	5	1
	12	2	-	1	-	1	1	4	2	3
	15	3	-	2	2	3	2	2	3	1
	1	4	-	-	-	-	1	-	-	-
	6	5	-	-	-	1	2	2	1	-
TOTAL:	54		-	4	2	6	11	15	11	5
8	16	1	1	2	2	2	5	2	2	-
	15	2	1	-	1	1	4	2	4	2
	6	3	-	-	1	1	2	1	1	-
	4	4	-	1	-	-	1	1	-	1
	11	5	1	-	1	2	3	2	2	-
TOTAL:	52		3	3	5	6	15	8	9	3

Distribution of Raven's Raw Scores by Grades and Groups.

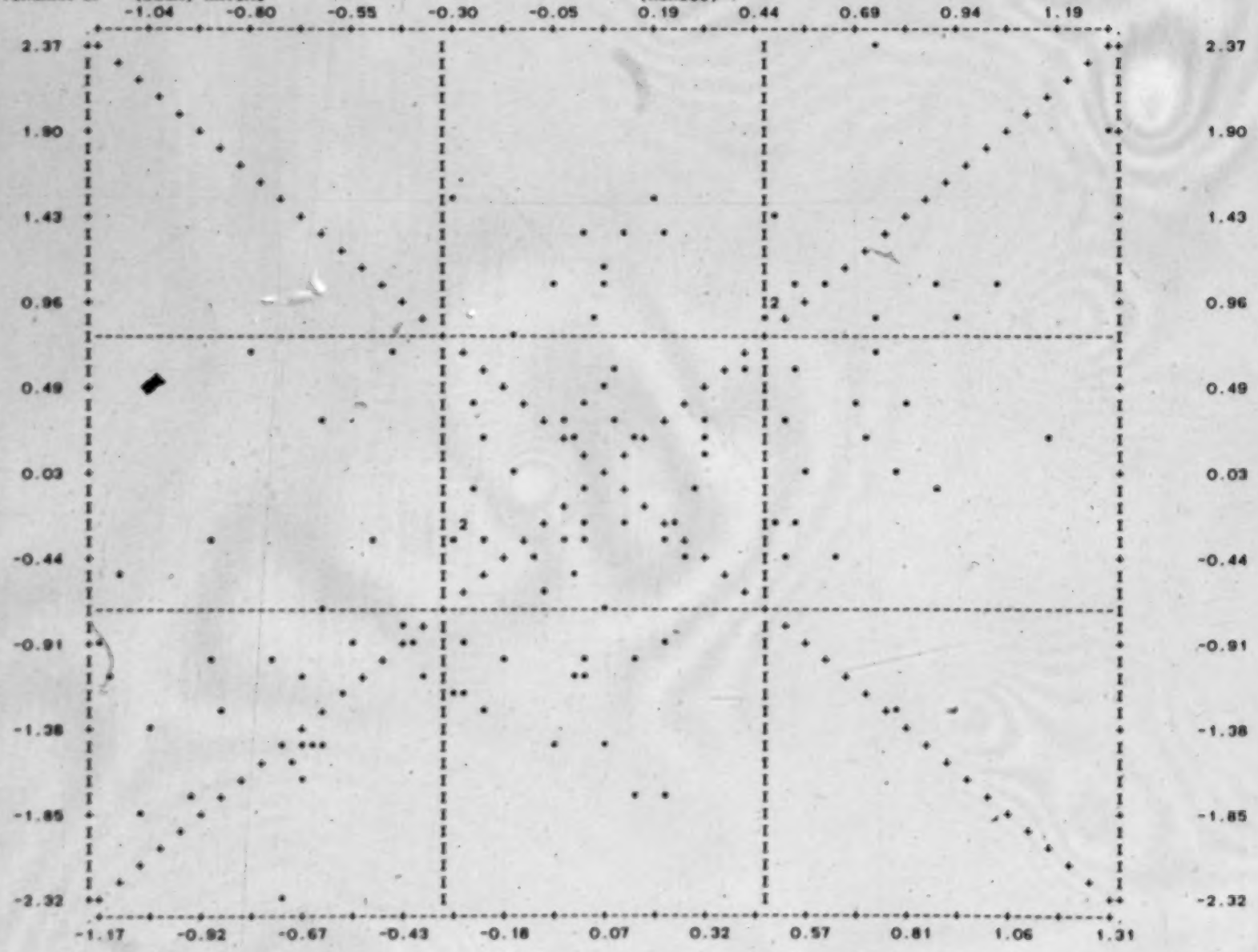
GRADE	N	GROUP	15	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50+
3	27	1	3	3	3	3	8	3	3	1	-
	10	2	-	-	-	1	1	4	4	-	-
	14	3	2	2	3	1	2	1	3	-	-
	7	4	-	1	2	1	-	1	2	-	-
	3	5	-	-	-	-	1	2	-	-	-
TOTAL:	61		5	6	8	6	12	11	12	1	-
4	18	1	-	2	1	1	2	5	4	2	1
	7	2	-	-	-	1	-	2	2	1	1
	18	3	2	1	-	3	5	2	2	2	1
	6	4	-	1	1	1	-	1	2	-	-
	6	5	-	-	-	1	1	-	1	1	2
TOTAL:	55		2	4	2	7	8	10	11	6	5
5	15	1	-	-	-	-	1	4	7	2	1
	8	2	-	-	1	-	-	-	3	3	1
	15	3	1	-	1	1	-	5	3	4	-
	8	4	-	-	2	-	2	1	1	1	1
	9	5	-	-	-	1	-	-	3	1	4
TOTAL:	55		1	-	4	2	3	10	17	11	7
6	15	1	-	1	-	-	1	3	4	4	2
	15	2	1	-	-	-	4	2	2	6	-
	18	3	1	-	-	1	1	-	8	3	4
	5	4	-	-	1	-	-	-	2	2	-
	2	5	-	-	-	-	-	1	-	1	-
TOTAL:	55		2	1	1	1	6	6	16	16	6
7	20	1	-	-	1	-	-	3	5	6	5
	12	2	-	-	-	-	1	2	3	5	1
	15	3	-	-	1	1	3	1	6	3	-
	1	4	-	-	-	-	-	-	-	1	-
	6	5	-	-	-	-	1	-	2	2	1
TOTAL:	54		-	-	2	1	5	6	16	17	7
8	16	1	-	-	-	2	1	1	3	7	2
	15	2	-	-	1	-	-	-	3	4	7
	6	3	-	-	-	-	1	-	2	3	-
	4	4	-	-	1	-	-	-	-	2	1
	11	5	-	-	-	1	-	-	2	4	4
TOTAL:	52		-	-	2	3	2	1	10	20	14

APPENDIX B

Scattergrams Indicating the Extent of Agreement Between the Raven's and L-T (FS) for Each Group and Total Group

SCATTERGRAM AND CROSSTABS FOR RAVENS, LORGE THORNDIKE 07/27/81 PAGE 2

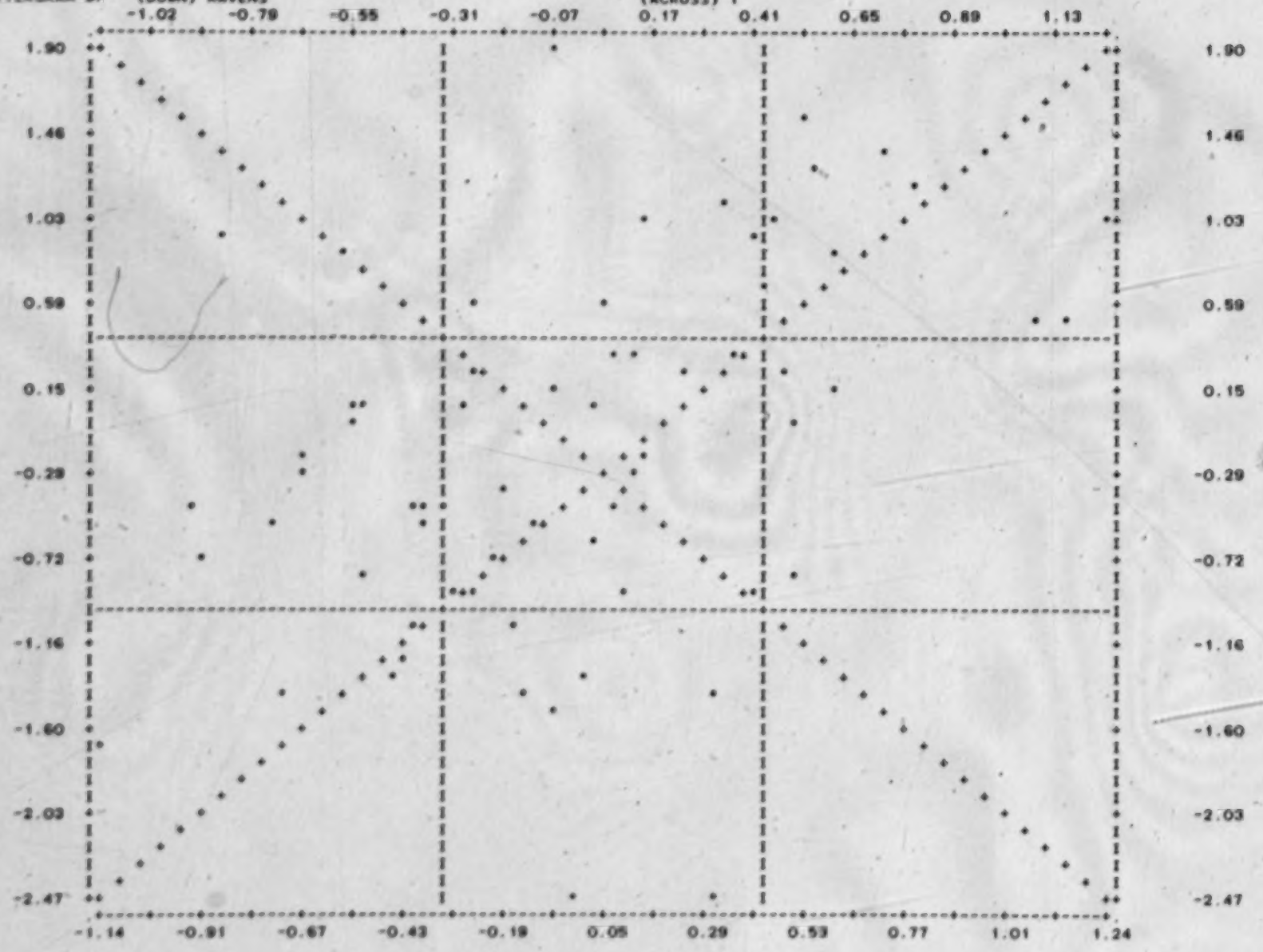
FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE ENG
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) T



University of Alberta

SCATTERGRAM AND CROSSTABS FOR RAVENS, LORGE THORNDIKE 07/27/81 PAGE 4

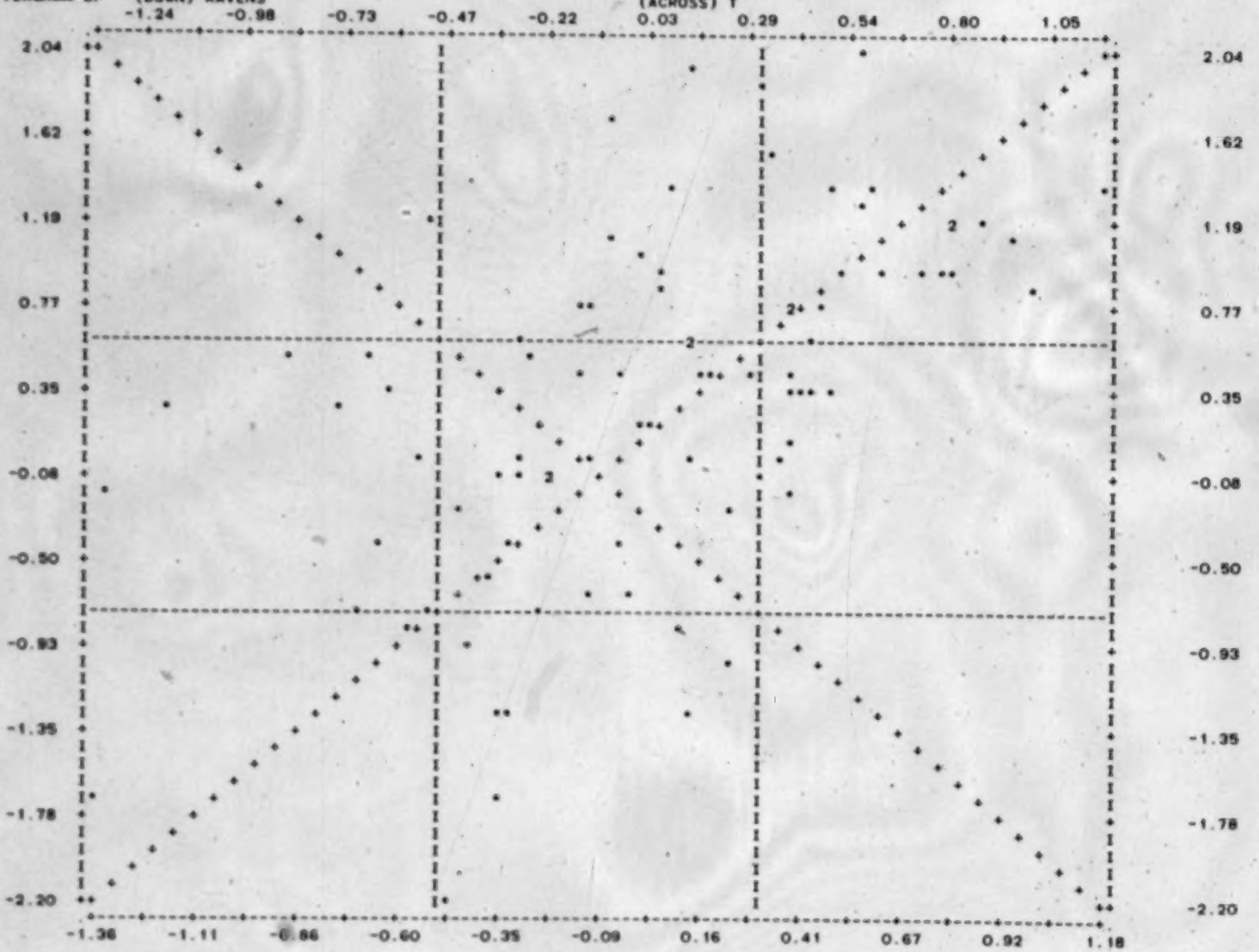
FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE FR
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) T



University of Alberta

SCATTERGRAM AND CROSSTABS FOR RAVENS, LORGE THORNDIKE 07/27/81 PAGE 6

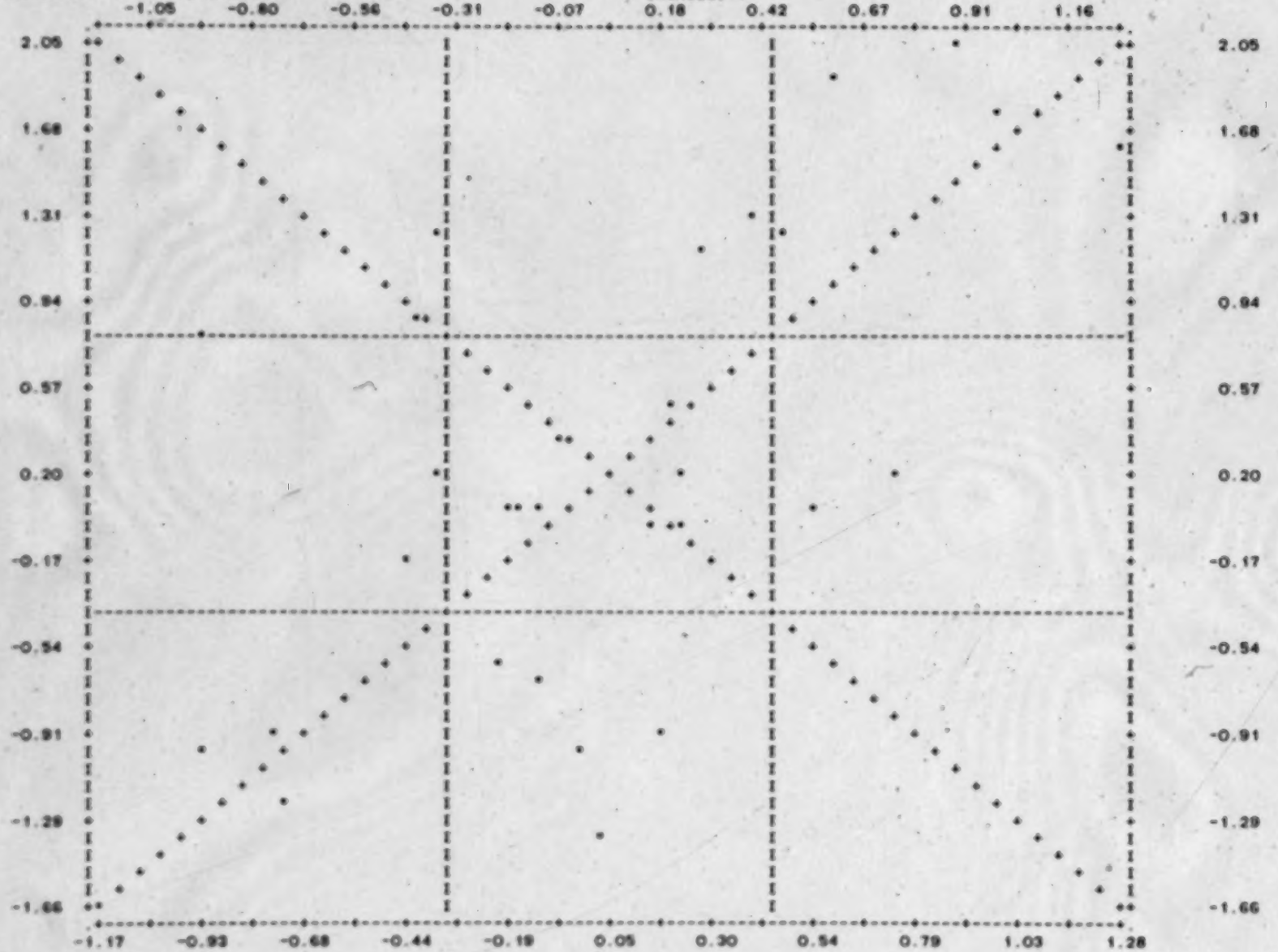
FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE CREE
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) T



University of Athens

SCATTERGRAM AND CROSSTABS FOR RAVENS, LORGE THORNDIKE 07/27/81 PAGE 8

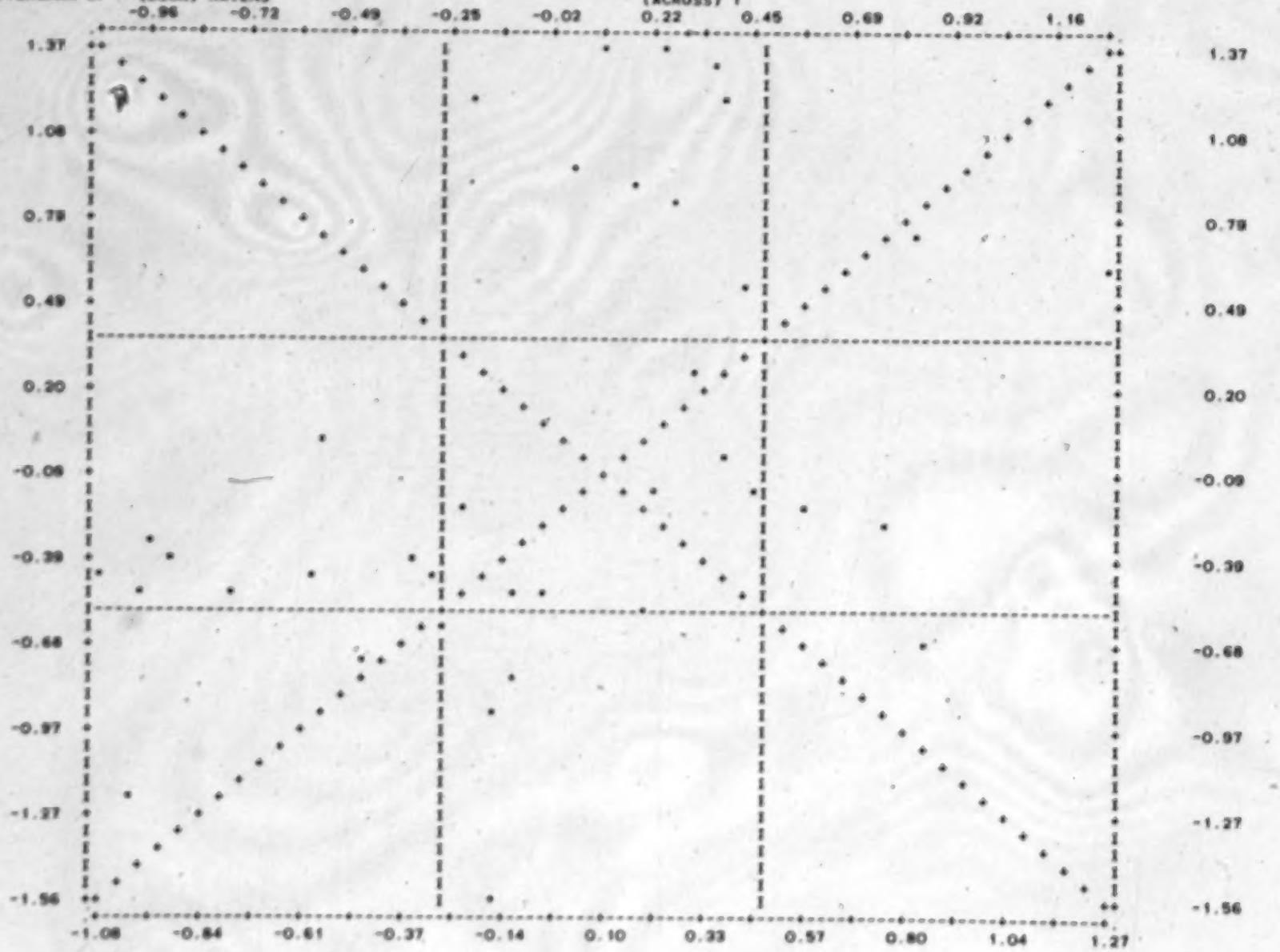
FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE CHIP
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) T



University of Alberta

SCATTERGRAM AND CROSSTABS FOR RAVENS, LORGE THORNDIKE 07/27/81 PAGE 10

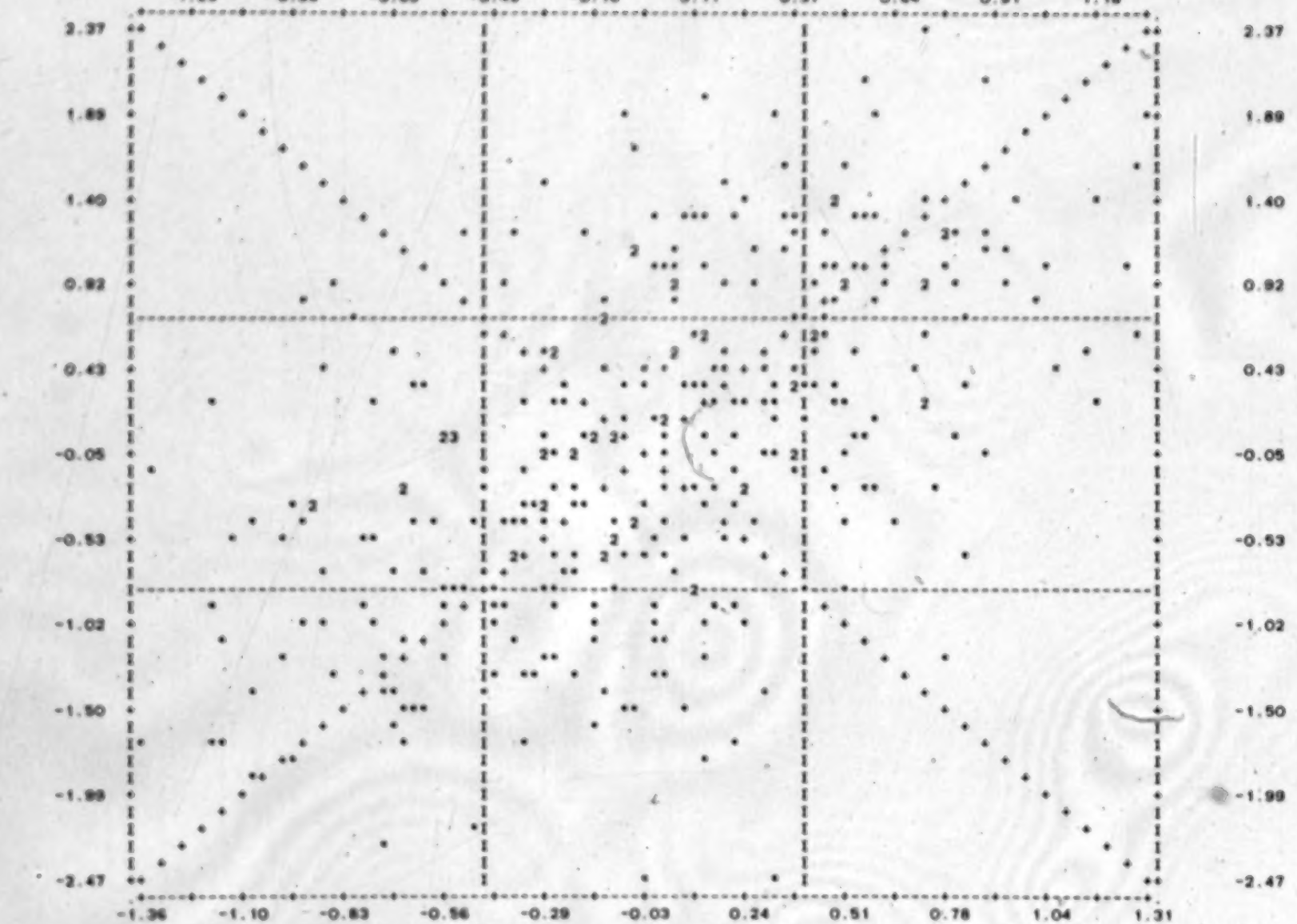
FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE OTHR
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) T



University of Athens

SCATTERGRAM AND CROSSTABS FOR RAVENS, LORGE THORNDIKE 07/27/81 PAGE 19

FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE ENG FR CREE CHIP OTHR
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) T



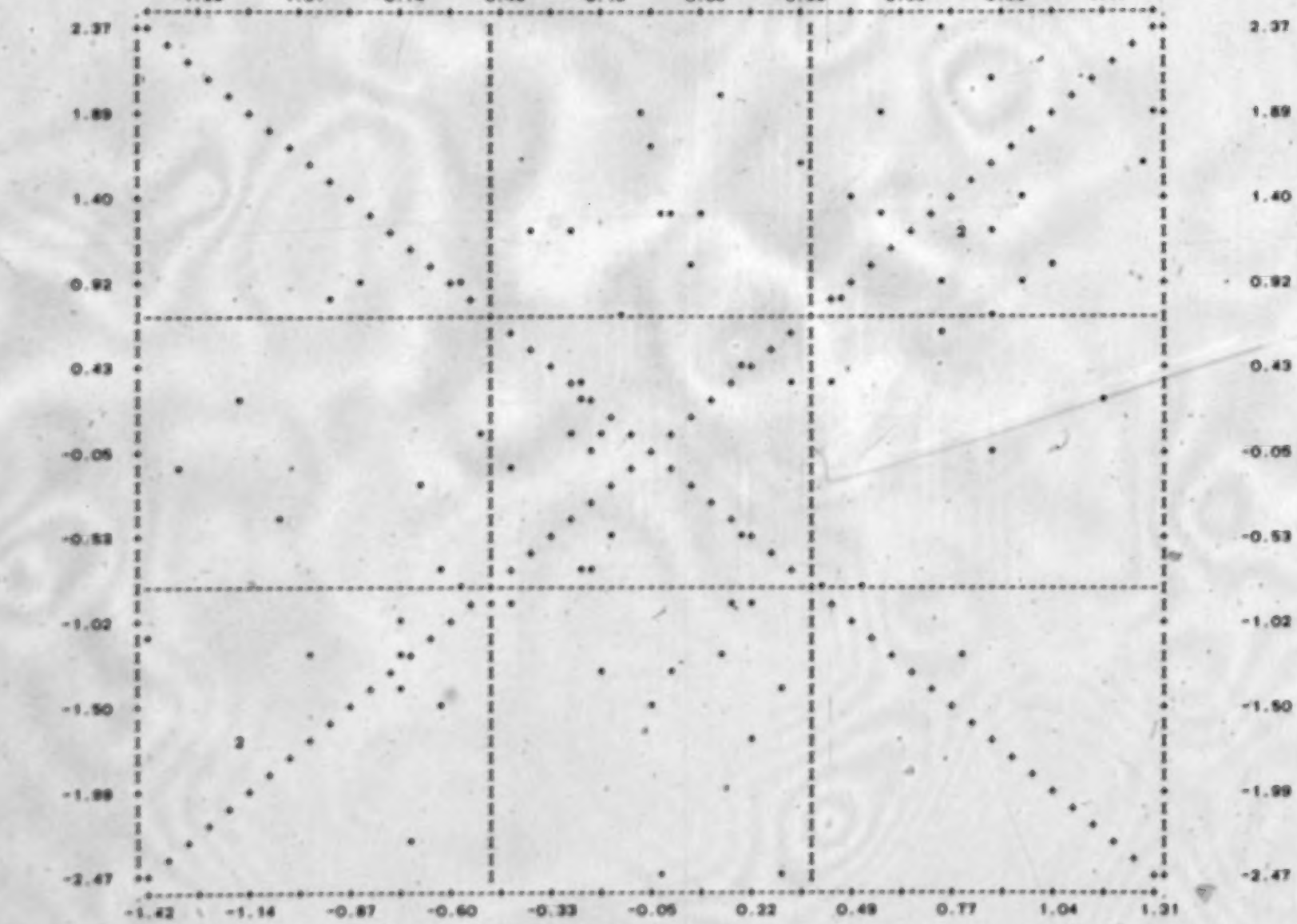
University of Athens

APPENDIX C

Scattergams Indicating the Extent of Agreement Among the Raven's, L-T (FS), and WISC-R (FS) for Total Group

SCATTERGRAMS FOR RAVENS, LORGE THORNDIKE AND WISCR 07/27/81 PAGE 2

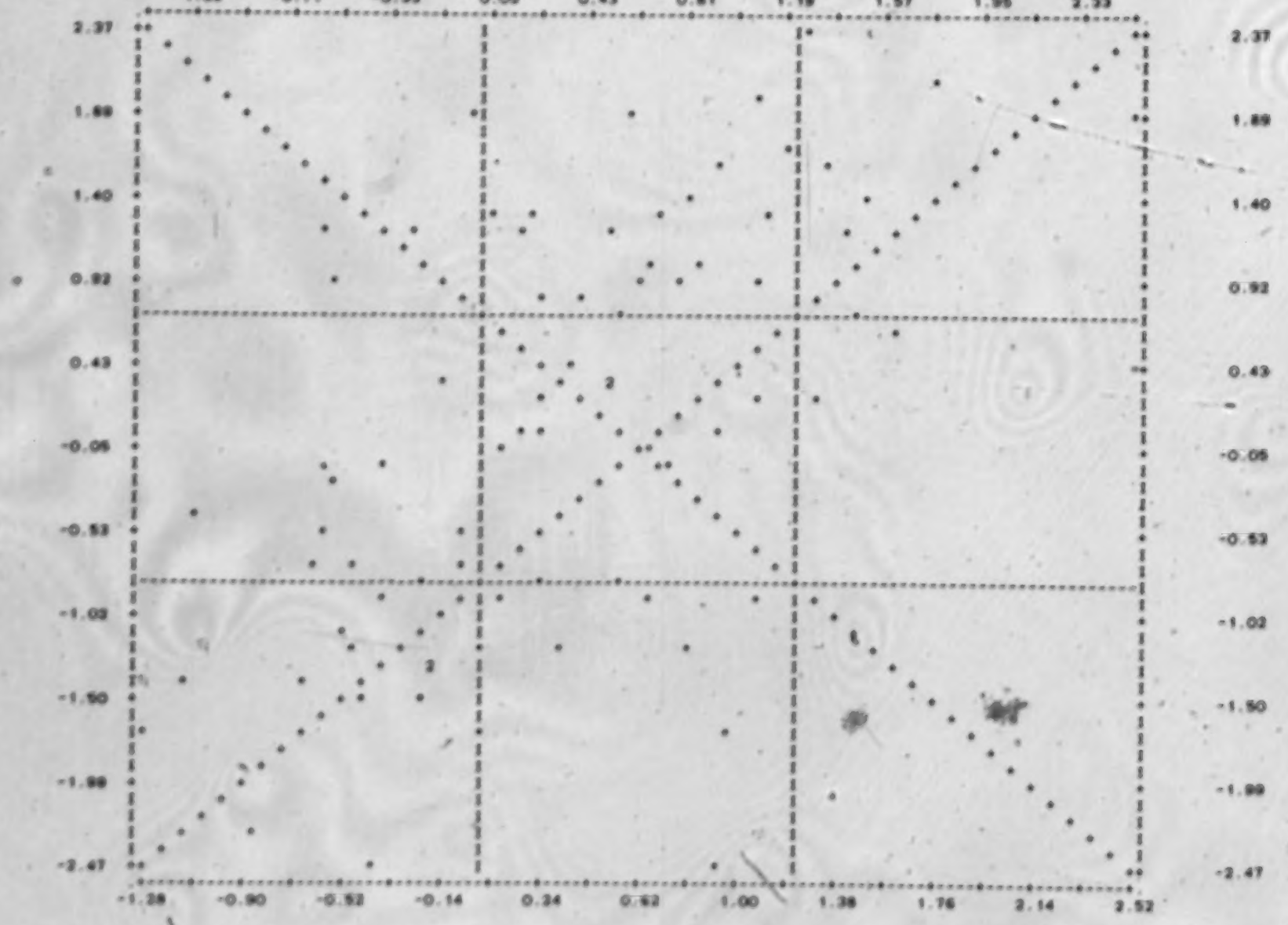
FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE ENG FR CREE CHIP OTHR
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) T



University of Alberta

SCATTERGRAMS FOR RAVENS, LORGE THORNDIKE AND WISCR 07/27/81 PAGE 4

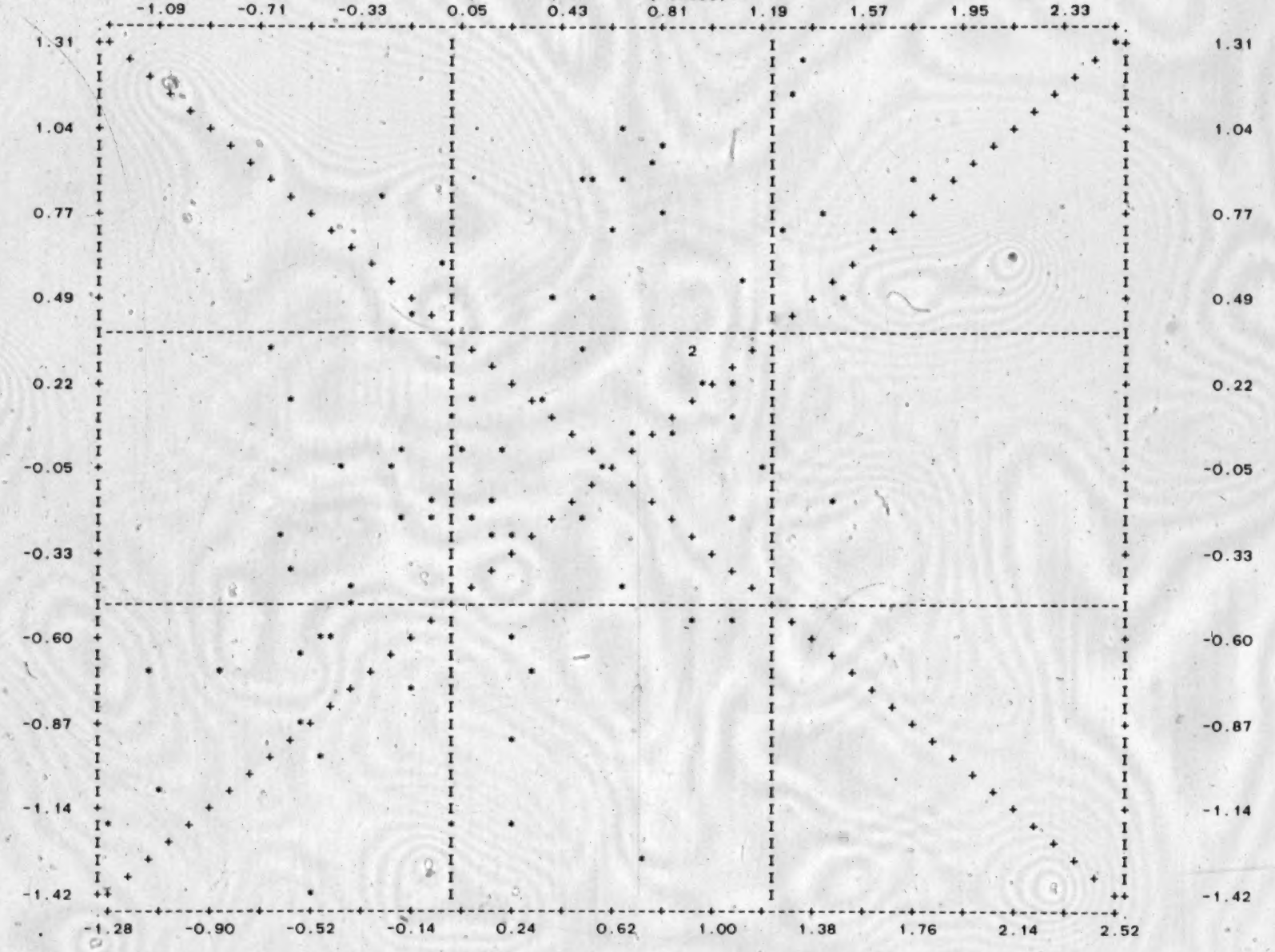
FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE ENG FB CREE CHIP OTHR
SCATTERGRAM OF (DOWN) RAVENS (ACROSS) W



University of Alberta

SCATTERGRAMS FOR RAVENS, LORGE THORNDIKE AND WISCR 07/27/81 PAGE 6

FILE NONAME (CREATION DATE = 07/27/81)
SUBFILE ENG FR CREE CHIP OTHR
SCATTERGRAM OF (DOWN) T (ACROSS) W



University of Alberta